EXPLORING MICROBIAL SMALL-MOLECULE CHEMISTRY USING 2D NMR- AND LC/ESI-MS-BASED COMPARATIVE METABOLOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ry Roger Forseth

January 2013

EXPLORING MICROBIAL SMALL-MOLECULE CHEMISTRY USING 2D NMR- AND LC/ESI-MS-BASED COMPARATIVE METABOLOMICS

Ry Roger Forseth, Ph.D.

Cornell University 2013

Microbial biology is integrated with a seemingly endless number of pathways involving small organic molecules. The coordinated regulation of biosynthesis, export, and detection of metabolites confer fitness for the producing microbes in nearly every environment. Continued, rapid evolution has yielded small-molecule biosynthetic enzymes capable of catalyzing an array of reactions that produce chemical species of high structural complexity, diversity, and target specificity.

The procurement rate of new chemical entities (NCEs) over the recent decades has decreased, largely due to the redundant discovery of previously characterized chemical species and an incomplete understanding of the regulation and biosynthetic steps involved in microbial metabolic pathways. Methodology to establish the full diversity of metabolites produced by a particular microbial species, its metabolome, and associated biosynthetic genes will greatly accelerate our ability to procure NCEs and explore their biological roles.

Metabolomics requires dedicated tools to characterize the small molecule products and biosynthetic potential of an organism. NMR spectroscopy and mass spectrometry are powerful analytical tools for unambiguously characterizing the structures of organic molecules. Therefore, the development of these methods for exploring microbial metabolism should foster access to the yet to be described metabolites of bacterial and fungal small molecule chemistry.

Described herein is the development of comparative metabolomics approaches, combining differential analysis by 2D NMR spectroscopy (DANS), LC/ESI-MS profiling, with specific gene overexpression, knock-out, or knock-down strategies to associate and characterize secondary metabolites with their corresponding biosynthetic genes or gene clusters.

Applied to the fungus *Aspergillus fumigatus's gli* cluster, DANS has revealed nine novel *gliZ*-dependent diketopiperazines, and their identification provides insight into gliotoxin biosynthesis. In a project involving the bacterium *Streptomyces clavuligerus*'s *hlm* gene cluster, comparative NMR and LC/ESI$^+$-MS demonstrated that the organism largely relies on alkylation though dimerization and/or methylation to detoxify deleteriously reactive thiol *hlm* pathway intermediates. Finally, as an extension to the *A. fumigatus* study, DANS is demonstrated to be a powerful tool for systematically examining the products of orphan secondary metabolic gene clusters, supporting the discovery of seven novel alkaloids in *Aspergillus flavus*.

BIOGRAPHICAL SKETCH

Ry Roger Forseth was born in Stevens Point, Wisconsin to Peter and Carol Forseth on November 5, 1984. He spent the majority of his adolescent life in Waupaca Wisconsin, attending Waupaca High School, from which he received his High School Diploma. Ry then went on to attended the University of Wisconsin-Madison, conducting organosilicon research with Professor Robert West, Department of Chemistry; and ecological water chemistry studies with Professor Emily Stanley, Center for Limnology. In May of 2007 Ry received a Bachelor of Science degree in chemistry from the College of Letters and Science. In the summer of 2007 Ry moved to Ithaca, New York and matriculated with the Department of Chemistry and Chemical Biology, Cornell University. Ry conducted his graduate research in the laboratory of Professor Frank C. Schroeder. Ry is very proud to be a survivor of testicular cancer, and is thankful for the support provided by the students, staff, faculty, and administrators of Cornell University during his battle with cancer.

Dedicated to: Sanjeev Vohra, M.D., his team at Ithaca Urology,

the medical professionals of

Cayuga Medical Center, and the

Melvin and Bren Simon Cancer Center, Indiana University.

Thank you for your continued fight against cancer.

ACKNOWLEDGMENTS

Special Committee Chair and Adviser: Professor Frank C. Schroeder.

Collaborating Principal Investigators: Professor Nancy P. Keller, Professor Christopher T. Walsh, Professor Dirk Hoffmeister, Professor Barbara J. Howlett, and Professor Albert A. Bowers.

Collaborating Researchers: Dr. Saori Amaike, Dr. Bo Li, Dr. Wei Jiang, Dr. Ellen M. Fox, DaWoon Chung, Daniel Schwenk, Katharyn J. Affeldt, Josh A. Baccile, Jenna A. Bernstein, and Yekaterina Rokhlenko

Special Committee: Professor Brian Crane, and Professor Hening Lin

Facility Managers: Dr. Ivan Keresztes, Dr. Marco Tonelli, Maciej Kukula, and Dave Kiemle

Organizations: Department of Chemistry and Chemical Biology and The Boyce Thompson Institute for Plant Research, Cornell University; Department of Medical Microbiology and Immunology and the National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School; SUNY College of Environmental Science and Forestry, Analytical and Technical Services.

Funding: Department of Chemistry and Chemical Biology Teaching Fellowship, Schroeder Laboratory Startup Funds, Chemistry Biology Interface Training Program (GM008500), Cornell University.

TABLE OF CONTENTS

PREFACE

Microbial Small-Molecule Chemistry: The biosynthetic transformations employed by microorganisms (bacteria, fungi, protozoa, and microscopic plants and animals) produce a vast range of macromolecular and small organic molecules. Exploring these anabolic pathways has greatly advanced agricultural,[1] nutritional,[2] medicinal,[3,4] and energy research,[5,6] and has significantly added to our understanding of microorganisms' influences in nature. Bacterial and fungal small molecule products (non-polymeric chemical species typically < 800 amu), often called natural products or secondary metabolites, have garnered significant interest from the chemical community, largely due to their often complex molecular architectures and potent biological activities.[3] These metabolites including terpenoids, polyketides, non-ribosomal peptides and mixed polyketide/non-ribosomal peptide hybrids are produced by a variety of biosynthetic enzymes, terpene cyclases, non-ribosomal peptide Synthetases (NRPSs), polyketide synthases (PKSs), and NRPS-PKS respectively, encoded for by secondary metabolic genes.[7] The structural characterization and biological activity assessment of several bacterial and fungal secondary metabolic products have led to the development of a number of small molecule therapeutics.[3]

Although there is significant interest in obtaining new chemical entities (NCE) from bacteria and fungi, the rate of small molecule discovery from these sources has slowed. The decrease in discovery has been attributed to the frequent rediscovery of previously characterize metabolites and an incomplete understanding of secondary metabolic gene regulatory pathways. Bioinformatics studies of the secondary metabolic gene content of fully and partially sequenced bacterial and fungal genomes have suggested the number of compounds described from microbial sources belies the biosynthetic potential these organisms process to produce NCE.[8] Additionally, secondary metabolic gene expression studies have demonstrate that laboratory culture conditions are often not sufficient to induce transcription of secondary metabolic genes,

indicating that these pathways likely require a specific set of chemical or physical signals to positively regulate secondary metabolic biosynthetic pathways.[9-12]

In many species of bacteria and fungi secondary metabolic gene disruption, through knockout or knockdown mutations, produces viable strains unable to biosynthesize individual or sets of secondary metabolites.[7] This attribute is advantageous for studying the metabolic changes that occur when specific biosynthetic pathways are disrupted. However, because these molecular species often play indispensable roles in defense,[13] virulence,[14] signaling,[15] pigmentation,[16] and/or reproduction, disruption of their biosynthesis has also been shown to significantly retard microbial survivability in stressful environmental conditions.[17,18] This has been demonstrated in studies conducted by the Keller Laboratory, University of Wisconsin-Madison, showing that the Δ*laeA* strain of the fungus *Aspergillus fumigatus*, unable to produce secondary metabolites, is significantly less virulent than wild-type in murine pulmonary infection models.[17]

Bacteria and fungi employ a host of integrated regulatory pathways to coordinate secondary metabolite biosynthesis, helping to confer fitness in a variety of environments. A common feature proposed to contribute to the regulation of secondary metabolite biosynthesis is the genomic location and organization of secondary metabolic genes.[19] In bacteria and fungi secondary metabolic genes are typically clustered[20] (fungi occasionally deviate from this observation, and include scattered secondary metabolic genes that are part of a unified biosynthetic pathway). The secondary metabolic gene clustering phenomenon has been attributed in part to intra and inter-kingdom horizontal gene transfer.[21-24] This exchange of genetic information explains why some bacterial metabolites, or close analogs thereof, have been found in fungal species and *vice versa*. Chapter 3 of this dissertation highlights one such example where two piperazines (**Figure P.1** compounds **1** and **2**) discovered from the fungus *Aspergillus flavus* have close structural, and likely a related biosynthetic

pathway, as piperazinomycin produced by the bacterium *Streptoverticillium olivoreticuli* (**Figure P.1**).

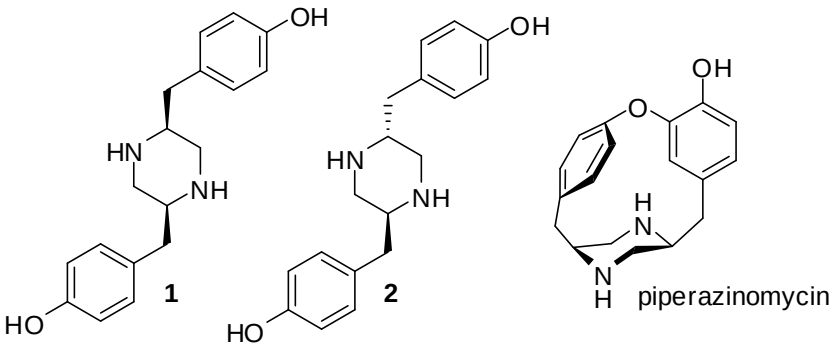


Figure P.1: Example for a structural motif present in both bacteria and fungi. Compounds **1** and **2** are products of a NRPS incorporating biosynthetic pathway used by the fungus *A. flavus*. Structurally similar compound, piperazinomycin, is a metabolite produced by *S. olivoreticuli*.

In addition to providing insight into the distribution and evolution of secondary metabolic genes,[20] the secondary metabolic gene clustering phenomenon has proven to be advantageous for defining the genetic components, and in turn regulatory, transport, and biosynthetic proteins involved in the production of a given secondary metabolite. The Walsh Laboratory, Harvard Medical School, has extensively utilized the clustering phenomenon of bacteria and fungi as an entry point to systematically examine the genes, and in turn proteins, involved in a given biosynthesis. Their approach often uses recombinant methodology to reconstitute active holo-enzymes *in vitro*, facilitating enzyme mechanistic and product identification studies. One of the major technological advances described by the Walsh Laboratory includes the development of *Bacillus subtilis*'s Sfp enzyme. Sfp is a promiscuous phosphopantetheinyl (PPT) transferase able to install an amino acid-functionalized PPT moiety onto an apo thiolation domain of a NRPS.[25,26] This technology effectively allows one to prepare holo-NRPS chemical species which can be studies *in vitro* along with proposed tailoring enzymes, likely encoded for in the same cluster as the holo-NRPS. (**Figure P.2**).

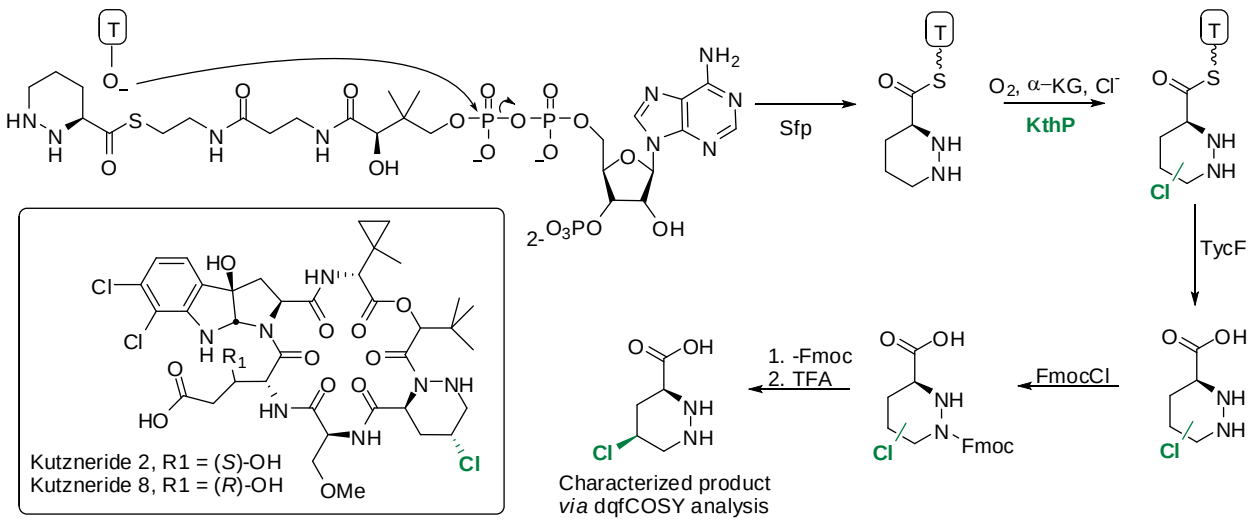


Figure P.2: Scheme depicting the implementation of *B. subtilis's* Sfp enzyme to study the enzyme KthP, and the structures of Kutznerides 2 and 8. Sfp was used to transfer the piperazate-functionalized PPT moiety to a free-standing thiolation domain. The holo-T domain was then studied in conjunction with KthP to determine the specific chemical reaction KthP catalyzes. Note that the stereo chemistry at C-5 of the characterized KthP product is *S* and in the natural products Kutznerides 2 and 8, that position in the piperazate moiety is *R*.

In collaborative work including the author of this dissertation and the Walsh Laboratory, the enzymatic product of the non-heme iron halogenase KthP was examined. KthP was hypothesized to chlorinate a piperazate substrate that, through further enzyme-mediated processing, is installed into the macrocyclic hexadepsipeptides kutzneride 2 and 8 from *Kutzneria sp.* 744 (**Figure P.2**).[27] This study joins a host of investigations aiming to define the scope of organic reactions accessed by bacteria and fungi to build complex small molecule scaffolds. However, as illustrated by the KthP study, these investigations can be quite laborious, and often have a narrow focus on one or two steps in an elaborate biosynthetic pathway. Additionally, because reconstituted *in vitro* biosynthetic pathways model complex *in vivo* systems that may contain yet to be defined factors modulating biosyntheses, complementary methods are needed to explore the corresponding endogenous biosynthetic pathways *in vivo*.[28] To

address this need, the work described in this dissertation develops comparative metabolomics tools to explore the metabolic changes that occur in bacteria and fungi in response to specific overexpression, knockout, and knockdown mutations made to secondary metabolic genes and gene clusters. These studies provide facile methodology to associated secondary metabolic genes with their corresponding metabolomic influence, and often yield information about the endogenous biological influence of theses pathways and metabolites.

Comparative Metabolomics, an Approach to Study Secondary Metabolism: NMR has traditionally been employed to unambiguously assign molecular structures for purified samples of organic molecules. Schroeder *et al.* published two ground breaking studies demonstrating that one can obtain full or partial structural information from 2D NMR spectra acquired for complex biological extracts containing mixtures of metabolites.[29,30] In these studies the structures of several reactive cyclic alkaloids produced by myrmicaria ants were obtained by utilizing the 2D NMR experiment, double quantum filter correlation spectroscopy (dqfCOSY), for crude ant extracts. dqfCOSY experiments provide ^{1}H-^{1}H coupling information including chemical shifts (diagnostic for the chemical environment of a proton in an organic molecule) and scalar *J*-coupling values (directly related to the dihedral angle between vicinal hydrogen atoms or the 2-bond coupling between germinal hydrogen atoms) useful values for assigning molecular structure.[30] Additionally, dqfCOSY experiments have a large dynamic range (>200 fold) able to capture correlations for a number of molecular species in a metabolic extract.

Analysis of metabolic mixtures using dqfCOSY took another technical step forward when it was applied to the characterization of the small molecule products associated with the orphan *pksX* gene cluster in the bacterium *B. subtilis*.[31] The secondary metabolites associated with the expression of *B. subtilis's pksX* gene cluster had largely remained uncharacterized due to their labile nature, preventing

chromatographic enrichment, a step commonly utilized in natural products characterization to isolate molecular species of interest. To overcome the detrimental chromatography step, a comparative metabolomics approach was utilized where crude extracts corresponding to a *pksX*$^{+}$ strain were compared to those derived from a *pksX*$^{-}$ strain. The comparison was carried out by acquiring dqfCOSY spectra for the *pksX*$^{+}$ and *pksX*$^{-}$ extracts, revealing the presence of cross peaks in the *pksX*$^{+}$ spectrum that were absent in the pksX^{-} spectrum. These *pksX*$^{+}$-specific signals supported the eventual characterization bacillaene a conjugated polyene containing an elimination-prone β-hydroxy amide.

Inspired by the comparative metabolomics approach utilized in the *B. subtilis* study, this dissertation extends differential analysis by 2D NMR spectroscopy (DANS) methodology to study the products, biosyntheses, and chemical biology of several bacterial and fungal secondary metabolic pathways (**Figure P.3**, DANS schematic). Additionally, parallel comparative LC/ESI-MS methods, provided an additional analytically orthogonal data set to enhance these studies.

Preview of Chapters: The chapters of this dissertation are reformatted and updated first or co-first authored publications composed by the author of this dissertation and publication co-authors. Each chapter includes the abstract, text, and figures previously published. Additional text and figures not originally included in the publications have been inserted to give a broader picture of the data acquisition, analysis and interpretation process. Where appropriate, updates have been included to highlight recently published work relating to the various chapter topics. Supporting information is included in the appendices found at the end of the dissertation. The research described herein is the result of collaborative efforts conducted between the author and several labs. Therefore, the descriptions given below highlight the content of each chapter and the contributions made by the individual authors of the original publication.

***Chapter 1,* Identification of Cryptic Products of the Gliotoxin Gene cluster Using NMR-Based Comparative Metabolomics and a Model for Gliotoxin Biosynthesis[32]:** Chapter 1 explores the metabolomic changes that result from knocking out *gliZ* (Δ*gliZ*), the transcriptional regulator of the gliotoxin gene cluster in *A. fumigatus*. The study employed DANS methodology (described in the previous section) to compare unfractionated, or partially enriched, extracts wild-type and Δ*gliZ* cultures (**Figure P.3**). These studies lead to the characterization of a family of *gliZ*-dependent diketopiperazines, including nine novel chemical species. The structural information provided by elucidation of the *gliZ*-dependent metabolites supported the development of a proposed gliotoxin biosynthetic scheme. This scheme addresses in detail the order and chemistry of the biosynthetic steps involved in gliotoxin production. These transformations are proposed as a collection of oxidative and addition reactions performed on a GliP (NRPS)-thioester tethered species, GliP-S-Ser-Phe-NH_2. This model stands in contrast to recently published work that suggest gliotoxin biosynthesis takes part though modification to untethered cyclo(Phe-Ser) species. As discussed at length at the end of Chapter 1, to date, both the tethered and untethered models seem to retain validity given the collection data that have been assembled pertaining to gliotoxin biosynthesis.

Authors' contributions: Ry R. Forseth, study design, data collection, data analysis and interpretation, figures, and writing; Ellen M. Fox and DaWoon Chung, study design, data collection, and data analysis pertaining to the Δ*gliI* mutant; Barbara J. Howlett, study design, data analysis, and writing; Nancy P. Keller, study design, data analysis and interpretation, and writing; and Frank C. Schroeder, study design, data analysis and interpretation, and writing.

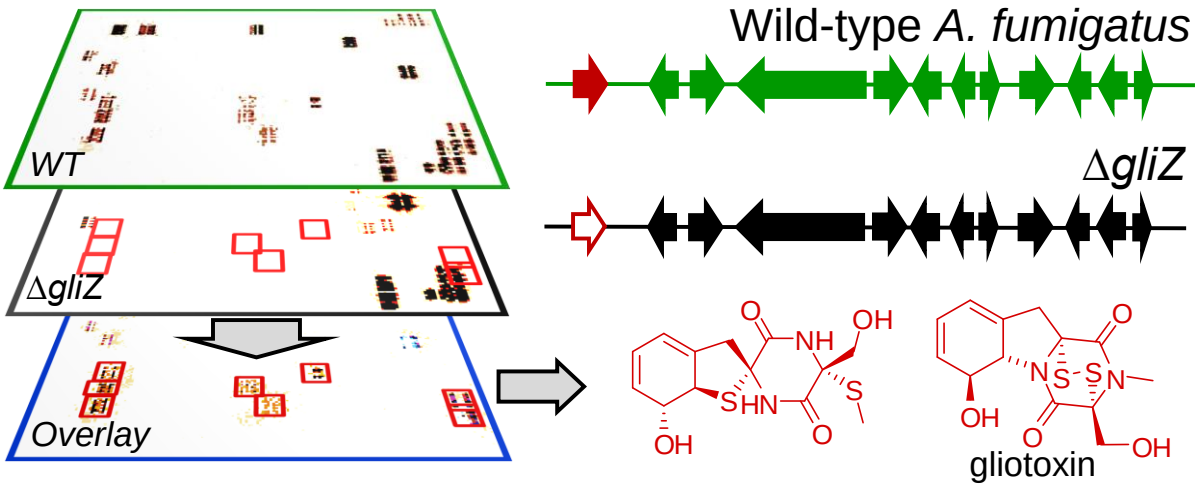


Figure P.3: Schematic overview of the DANS methodology as applied to the *A. fumigatus gliZ* study. Wild-type *A. fumigatus* expresses *gliZ*, the transcriptional regulator of the gliotoxin gene cluster, and in turn produces gliotoxin. In the Δ*gliZ* strain, a *gliZ* knockout mutant, expression of the gliotoxin gene cluster genes and production of associated metabolites is abolished. dqfCOSY spectroscopic comparison of extracts corresponding to wild-type and Δ*gliZ* reveals signals representing compounds present in wild-type and absent in Δ*gliZ.* Analysis of the wild-type unique signals provides a basis for structure elucidation of the *gliZ*-dependent metabolites.

***Chapter 2,* A Backup Plan for Self-Protection: *S*-Methylation of Holomycin Biosynthetic Intermediates in *Streptomyces clavuligerus*[33]:** Chapter 2 investigates the consequences of disrupting the ultimate or penultimate enzymatic steps in holomycin biosynthesis. The bacterium *S. clavuligerus*'s *hlm* gene cluster encodes the enzymes responsible for the production of a dithiolopyrrolone antibiotic holomycin (**Figure P.5**). The holomycin biosynthetic pathway utilizes NRPS-supported chemistry to condense two equivalents of cysteine, and oxidize/decarboxylate yet to be described intermediates. These steps lead to the production a dithiol analog of holomycin which was found to undergo HlmI-catalyzed oxidation to the corresponding disulfide. Further, HlmI is needed to confer producer resistance to holomycin toxicity. Unknown was the fate of homomycin intermediates produced by *S. clavuligerus* strains containing an *hlmI* knockout background (Δ*hlmI*). Comparative metabolomics using ^{1}H NMR, ^{1}H,^{13}C-HMBC

and comparative LC/ESI-MS revealed a number of *S*-methylated analogs of holomycin and a heterodimer species (**Figure P.5**). The results of these studies suggest that capping the free thiol with a methyl group and hetero dimerization represents two major transformation pathways utilized by *S. clavuligerus* to deal with the accumulation of reduced holomycin intermediates.

Authors' contributions: Bo Li and Ry R. Forseth contributed equally to this publication and conducted study design, data collection, data analysis and interpretation, figures, and writing; Frank C. Schroeder, study design, data analysis and interpretation and writing; and Christopher T. Walsh, study design, data analysis and interpretation and writing.

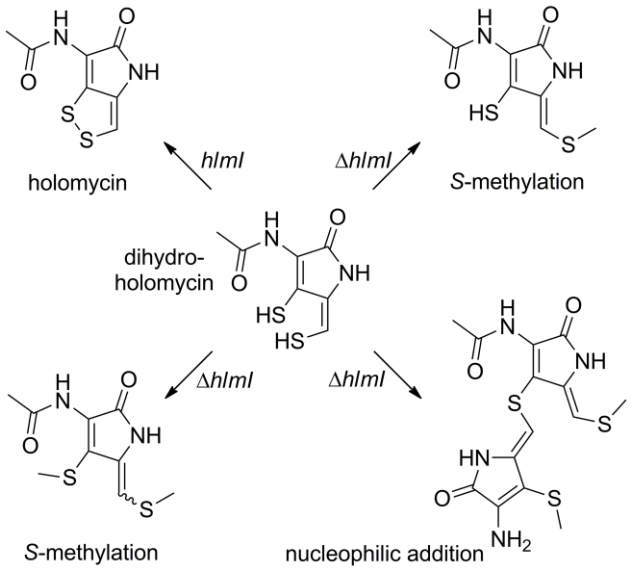


Figure P.4: Representation of the various small molecule transformation pathways utilized by *S. clavuligerus* to process free thiol-containing holomycin intermediates. In the *hlmI* background *S. clavuligerus* produces holomycin. Disruption of the *hlmI* gene results in methylation of a common dithiol-containing intermediate that may be singly methylated, doubly methylated or form hetero dimeric species.

Chapter 3, **Homologous Non-Canonical NRPS Gene Clusters Mediate Redundant Small-Molecule Biosynthesis in *Aspergillus flavus* (in press):** Chapter 3 extends the DANS methodology described in the Chapter 1 *A. fumigatus gliZ* study by developing comparative gene knockout, knockdown, and overexpression approaches to probe small molecule outputs and biosynthetic pathways involving orphan secondary metabolic gene cluster (gene clusters for which there is no known associated metabolite). This chapter focuses on the orphan *lna* and *lnb* clusters from the fungus *A. flavus*. The *lna* and *lnb* clusters both contain one NRPS gene. Interestingly, the NRPS gene of the *lna* cluster (*lnaA*) is homologous to *lnb*'s NRPS gene (*lnbA*), and both NRPS genes are clustered with similar sets of genes encoding putative small molecule tailoring enzymes. Both *lnaA* and *lnbA* encode a three module NRPS, putatively able to activate, tether, and reduce one amino acid. To investigate the products associated with these NRPS genes, and in turn clusters, a series of *lnaA* knock out and overexpression *A. flavus* strains were constructed as well as several double mutants including a knockout *lnaA* and knockdown *lnbA* strain (Δ*lnaA, KD::lnbA*). DANS and comparative LC/ESI-MS of the various *A. flavus* strains revealed a family of alkaloids including piperazines-, pyrazines-, and morpholine-containing moieties. These species are derived from the coupling of two equivalents of L-tyrosine. Structural and mutant profiling information for the discovered *lna*/*lnb*-metabolites provided insight into their biosynthesis. Complementing the chemical investigations of this chapter are several important biochemical, phenotype, and expression studies confirming the substrate specificity of LnaA and LnbA and providing insight into the influence of the discovered metabolites on phenotype and gene expression (**Figure P.5**).

Authors' contributions: Ry R. Forseth and Saori Amaike contributed equally to this publication and conducted study design, data collection, data analysis and interpretation, figures, and writing; Daniel Schwenk, LnaA and LnbA biochemical study design, data collection, and data interpretation; Katharyn J. Affeldt, gene

expression data collection, data analysis, and figures; Dirk Hoffmeister, study design, data analysis and interpretation, and writing; Frank C. Schroeder, study design, data analysis and interpretation and writing; Nancy P. Keller, study design, data analysis and interpretation and writing.

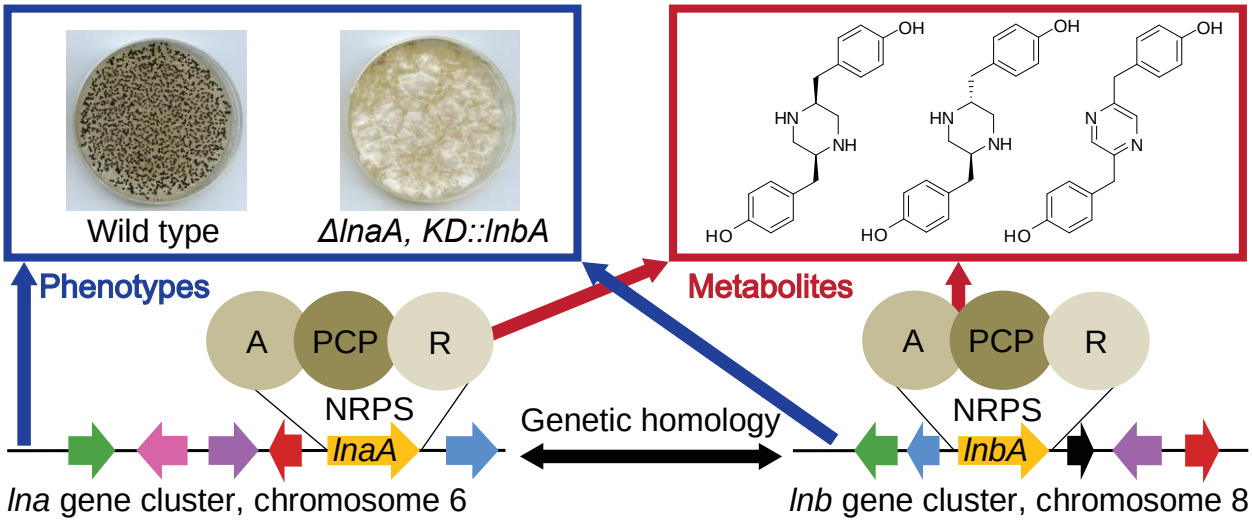


Figure P.5: Depiction of results from the *A. flavus* study. This investigation presents a comparative metabolomics approach to study *A.s flavus's* orphan *lna* and *lnb* secondary metabolite gene clusters. Interestingly, these two clusters encode partially redundant biosynthesis pathways, responsible for the production of a series of novel fungal alkaloids (red box) and appear to be integrated in the regulation of sclerotial formation (blue box), an important survival body produced by *A. flavus*.

REFERENCES

(1) Baker, B.; Zambryski, P.; Staskawicz, B.; Dinesh-Kumar, S. P. *Science* **1997**, *276*, 726.

(2) Lozupone, C. A.; Stombaugh, J. I.; Gordon, J. I.; Jansson, J. K.; Knight, R. *Nature* **2012**, *489*, 220.

(3) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2012**, *75*, 311.

(4) Walsh, C. *Antibiotics : actions, origins, resistance*; ASM Press: Washington, D.C., **2003**.

(5) Logan, B. E.; Rabaey, K. *Science* **2012**, *337*, 686.

(6) Hasunuma, T.; Okazaki, F.; Okai, N.; Hara, K. Y.; Ishii, J.; Kondo, A. *Bioresource technology* **2012**.

(7) Roze, L. V.; Chanda, A.; Linz, J. E. *Fungal Genet Biol* **2011**, *48*, 35.

(8) Hertweck, C. *Nat Chem Biol* **2009**, *5*, 450.

(9) Lim, F. Y.; Sanchez, J. F.; Wang, C. C.; Keller, N. P. *Methods in enzymology* **2012**, *517*, 303.

(10) Winter, J. M.; Behnken, S.; Hertweck, C. *Current Opinion in Chemical Biology* **2011**, *15*, 22.

(11) Brakhage, A. A.; Schroeckh, V. *Fungal Genet Biol* **2011**, *48*, 15.

(12) Georgianna, D. R.; Fedorova, N. D.; Burroughs, J. L.; Dolezal, A. L.; Bok, J. W.; Horowitz-Brown, S.; Woloshuk, C. P.; Yu, J.; Keller, N. P.; Payne, G. A. *Mol Plant Pathol* **2010**, *11*, 213.

(13) Oh, D. C.; Poulsen, M.; Currie, C. R.; Clardy, J. *Nat Chem Biol* **2009**, *5*, 391.

(14) Sugui, J. A.; Pardo, J.; Chang, Y. C.; Zarember, K. A.; Nardone, G.; Galvez, E. M.; Mullbacher, A.; Gallin, J. I.; Simon, M. M.; Kwon-Chung, K. J. *Eukaryot Cell* **2007**, *6*, 1562.

(15) Parsek, M. R.; Greenberg, E. P. *PNAS USA* **2000**, *97*, 8789.

(16) Ishida, K.; Lincke, T.; Behnken, S.; Hertweck, C. *J Am Chem Soc* **2010**, *132*, 13966.

(17) Bok, J. W.; Balajee, S. A.; Marr, K. A.; Andes, D.; Nielsen, K. F.; Frisvad, J. C.; Keller, N. P. *Eukaryot Cell* **2005**, *4*, 1574.

(18) Kale, S. P.; Milde, L.; Trapp, M. K.; Frisvad, J. C.; Keller, N. P.; Bok, J. W. *Fungal Genet Biol* **2008**, *45*, 1422.

(19) Hoffmeister, D.; Keller, N. P. *Nat Prod Rep* **2007**, *24*, 393.

(20) Osbourn, A. *Trends in genetics : TIG* **2010**, *26*, 449.

(21) Ginolhac, A.; Jarrin, C.; Robe, P.; Perriere, G.; Vogel, T. M.; Simonet, P.; Nalin, R. *J Mol Evol* **2005**, *60*, 716.

(22) Wenzl, P.; Wong, L.; Kwang-won, K.; Jefferson, R. A. *Mol Biol Evol* **2005**, *22*, 308.

(23) Marcet-Houben, M.; Gabaldon, T. *Trends in Genetics : TIG* **2010**, *26*, 5.

(24) Schmitt, I.; Lumbsch, H. T. *PloS one* **2009**, *4*, e4437.

(25) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585.

(26) Belshaw, P. J.; Walsh, C. T.; Stachelhaus, T. *Science* **1999**, *284*, 486.

(27) Jiang, W.; Heemstra, J. R., Jr.; Forseth, R. R.; Neumann, C. S.; Manaviazar, S.; Schroeder, F. C.; Hale, K. J.; Walsh, C. T. *Biochemistry* **2011**, *50*, 6063.

(28) Forseth, R. R.; Schroeder, F. C. *Current Opinion in Chemical Biology* **2011**, *15*, 38.

(29) Schroder, F.; Franke, S.; Francke, W.; Baumann, H.; Kaib, M.; Pasteels, J. M.; Daloze, D. *Tetrahedron* **1996**, *52*, 13539.

(30) Schroder, F.; Sinnwell, V.; Baumann, H.; Kaib, M. *Chem Commun* **1996**, 2139.

(31) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straightt, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *PNAS USA* **2007**, *104*, 1506.

(32) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(33) Li, B.; Forseth, R. R.; Bowers, A. A.; Schroeder, F. C.; Walsh, C. T. *Chembiochem* **2012**, *13*, 2521.

CHAPTER 1

IDENTIFICATION OF CRYPTIC PRODUCTS OF THE GLIOTOXIN GENE CLUSTER USING NMR-BASED COMPARATIVE METABOLOMICS AND A MODEL FOR GLIOTOXIN BIOSYNTHESIS

Abstract: Gliotoxin, a major product of the *gli* non-ribosomal peptide synthetase gene cluster, is strongly associated with virulence of the opportunistic human pathogen *Aspergillus fumigatus*. Despite identification of the *gli* cluster, the pathway of gliotoxin biosynthesis has remained elusive, in part because few potential intermediates have been identified. In addition, previous studies suggest that knowledge of *gli*-dependent metabolites is incomplete. Here differential analysis by 2D NMR spectroscopy (DANS) of metabolite extracts derived from *gli* knock-out and wild-type strains is used to obtain a detailed inventory of *gli*-dependent metabolites. DANS-based comparison of the wild-type metabolome with that of *ΔgliZ*, a knock-out strain devoid of the gene encoding the transcriptional regulator of the *gli* cluster, revealed nine novel *gliZ*-dependent metabolites including unexpected structural motifs. Their identification provides insight into gliotoxin biosynthesis and may benefit studies of the role of the *gli* cluster in *A. fumigates* virulence. This study demonstrates the utility of DANS for correlating gene expression and metabolite biosynthesis in microorganisms.

Introduction: Filamentous fungi produce remarkably diverse metabolomes including many small molecules of polyketide synthase (PKS) or non-ribosomal peptide synthetase (NRPS) origin that play important roles in pathogenesis.[1,2] The opportunistic pathogen *Aspergillus fumigatus*, a causative agent of invasive aspergillosis, produces copious amounts of gliotoxin (**1**, **Figure 1.1**), a representative member of a small family of epipolythiodioxopiperazines (ETPs) that is strongly associated with *A. fumigatus*

virulence.[3,4] Gliotoxin and related ETPs are products of the *gli* NRPS gene cluster.[5] Knock-out mutations of *gliP* (*ΔgliP*), a three-module NRPS (see Appendix A, **Figure A.1** for domain architecture),[5,6] *gliI* (*ΔgliI*), encoding a putative pyridoxal 5'-phosphate (PLP) binding domain,[7,8] or *gliZ* (*ΔgliZ*), a Zn_2Cys_6 binuclear transcription factor, abolishes gliotoxin biosynthesis.[9-12] *ΔgliP* has been shown to be significantly less virulent than the wild-type strain in immunosuppressed mice,[11] confirming that gliotoxin and perhaps other *gli*-dependent metabolites play a role in overcoming host resistance.

Previous studies suggest that knowledge of *gli*-dependent metabolites is incomplete.[12] In addition, the sequence of steps in the biosynthesis of the gliotoxins has remained unclear, in part because few potential biosynthetic intermediates or shunt metabolites have been identified. In this study it was hypothesized that NMR-based comparative metabolomics of *gli*-knock-out and wild-type or *gli*-overexpressing strains could provide a comprehensive overview of *gli*-dependent metabolites, including shunt metabolites and other cryptic products.[13,14] Recent studies, including the identification of bacillaene as the product of the mixed PKS-NRPS gene cluster *pksX* in *Bacillus subtilis*[15] and the identification of mating pheromones in *Caenorhabditis elegans*,[16] have demonstrated the utility of differential analysis by 2D NMR spectroscopy (DANS) for connecting metabolites with their biosynthetic pathways. In these examples, DANS combined with HPLC-MS analyses provided a comprehensive overview of the metabolic changes caused by knocking out small-molecule biosynthetic genes, which in each case led to the identification of several previously undetected metabolites derived from the knocked-out pathway.

DANS Comparative Metabolic Analysis, ΔgliZ and Wild-Type: For metabolic comparison, wild-type *A. fumigatus* and the mutant strain *ΔgliZ* were used.[12] Knock-outs of *gliZ* have been shown to stop the expression of the majority of genes in the *gli* cluster with the exception of *gliT*.[12,17,18] Therefore, DANS-based comparison of the *ΔgliZ* and

wild-type metabolomes should enable identification of any metabolites whose biosyntheses directly or indirectly depend on *gli* expression (**Figure 1.1**).

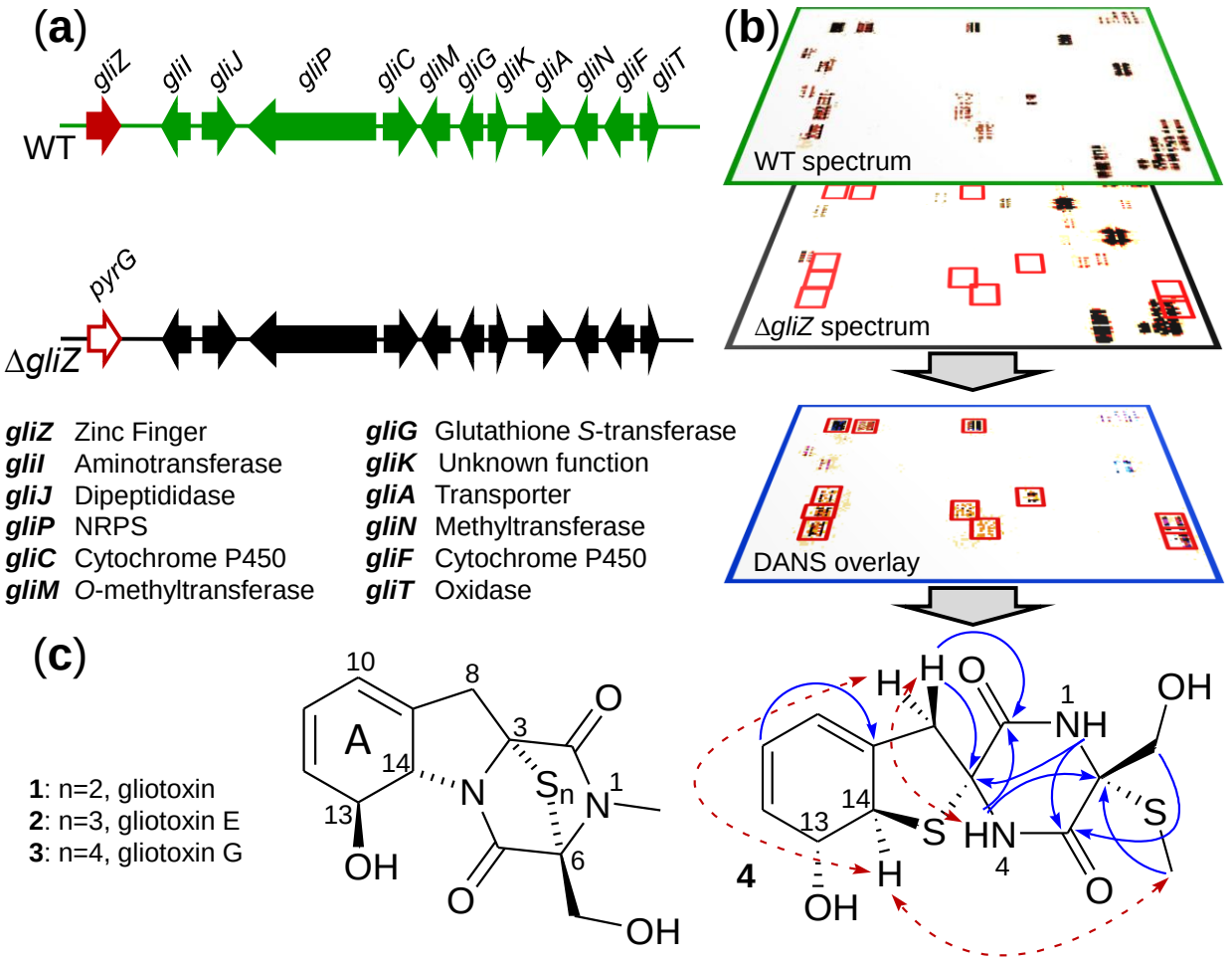


Figure 1.1: Identification of *gliZ*-dependent metabolites *via* DANS. (**a**) Gliotoxin gene cluster in *A. fumigatus* wild-type and mutant strain *ΔgliZ*, in which *gliZ* is replaced by the selection marker *pyrG*, encoding orotidine-5'-phosphate decarboxylase. (**b**) DANS overlay technique (schematic). Signals in the DANS overlay serve as markers for compounds whose biosyntheses depend on *gliZ* expression. (**c**) Examples for *gliZ*-dependent metabolites identified in this study and structure elucidation of **4** (^{1}H,^{13}C-HMBC, blue arrows; ROESY, red arrows).

In preparation for NMR spectroscopic analyses, *ΔgliZ* and wild-type metabolic extracts were fractionated into three metabolite pools of reduced complexity and limited polarity range, which ensured that both very polar (pool 1) and very nonpolar (pool 3)

metabolites would be detected (see Appendix A, section 5) For the resulting three wild-type pools and three *ΔgliZ* pools, dqfCOSY spectra were acquired. Compared to other 2D NMR spectroscopic techniques such as TOCSY[19] or HSQC,[20] dqfCOSY offers distinct advantages for the detection of novel or unanticipated compounds, because cross peak fine structure in dqfCOSY spectra provides greater structural information, including full signal multiplicity and coupling constants.[21] Furthermore, dqfCOSY cross peak fine structures can be easily modeled,[21] which helps resolve peak overlap and facilitates recognition of minor components.[16] dqfCOSY spectra also offer better sensitivity than HSQC, often enabling characterization of trace components representing < 0.1% of a sample.[22,23]

For DANS, dqfCOSY spectra obtained for the three wild-type metabolite pools were compared with the corresponding *ΔgliZ* pools using an overlay algorithm designed to highlight wild-type signals that were completely absent from the *ΔgliZ* spectra. This approach excluded compounds from the analysis whose biosynthesis was not entirely *gliZ*-dependent. DANS revealed more than 20 distinct *gliZ*-dependent spin systems in wild-type pool 2 (**Figures 1.2** and Appendix A, **Figure A.2** and **Figure A.3**) and a smaller number of *gliZ*-dependent signals in wild-type pool 1. No *gliZ*-dependent signals were observed in wild-type pool 3.

Characterization of gliZ-Dependent Metabolites' Structures: DANS analysis of pool 1 revealed two major *gliZ*-dependent spin systems that were accompanied by two closely related sets of less abundant signals. These spins systems suggested the presence of two structurally related *gliZ*-dependent compounds that feature a 5,6-disubstituted cyclohexa-2,4-dien-1-ol bearing a methylene group in position-5 (C-8, **4**) analogous to the A-ring of the gliotoxins (**Figure 1.1**). However, the dqfCOSY spectrum revealed an unusually large coupling constant of 16 Hz between the two methine protons in the A-ring moiety of the pool 1 analytes, which is significantly larger than the

coupling constants between the corresponding protons of gliotoxin (13-14 Hz). Initial attempts to isolate or enrich the major *gliZ*-dependent metabolite in pool 1 *via* preparative HPLC resulted in almost complete loss of the compound, indicating limited chemical stability. Therefore, a larger sample of pool 1 was prepared and fractionated using carefully conditioned flash chromatography media (Appendix A, section 8). Chromatographic enrichment of the **4** was monitored using diagnostic ^{1}H NMR signals established in the pool 1 DANS analysis. dqfCOSY spectra acquired for the purified metabolite confirmed the DANS-based assignments of the ^{1}H spin systems, and in conjunction with ^{1}H,^{13}C-HMBC and ^{1}H,^{13}C-HMQC data revealed that this compound represented a diketopiperazine based on the same serine- and phenylalanine-derived building blocks as the gliotoxins (**1**-**3**). Despite these similarities, the spectroscopic data suggested the presence of significant structural differences in the phenylalanine-derived portion of **4**. The chemical shift of C-3 (69.6 ppm) suggested sulfur-substitution at this position. However, the chemical shift of C-14 (57.1 ppm) was lower than in the gliotoxins, which carry an amide nitrogen in this position. This suggested that C-14 in **4** may not be attached to nitrogen, which is consistent with the observation of two amide protons in the NMR spectra. High resolution mass spectrometry (Appendix A, section **10**) indicated a molecular formula of $C_{13}H_{16}N_2O_4S_2$ which, in conjunction with ROESY and ^{1}H,^{13}C-HMBC data, established the structure of **4** as shown, featuring a sulfur bridge between C-3 and C-14 as part of a 6,9-diaza-1-sulfaspiro[4.5]decane system. ROESY spectra showed correlations between the protons of the methylsulfanyl attached to C-6 and the proton attached to C-14 supporting the relative configuration of **4**. Compound **4** is a novel metabolite with a structural motif rarely observed in nature; a sulfur- and nitrogen-bound spiro atom. The second *gliZ*-dependent metabolite detected in pool 1 could not be fully identified due to its lower abundance and limited chemical stability. However, the available NMR spectroscopic data for the compound strongly suggest that it represents a stereoisomer of spiro compound **4**.

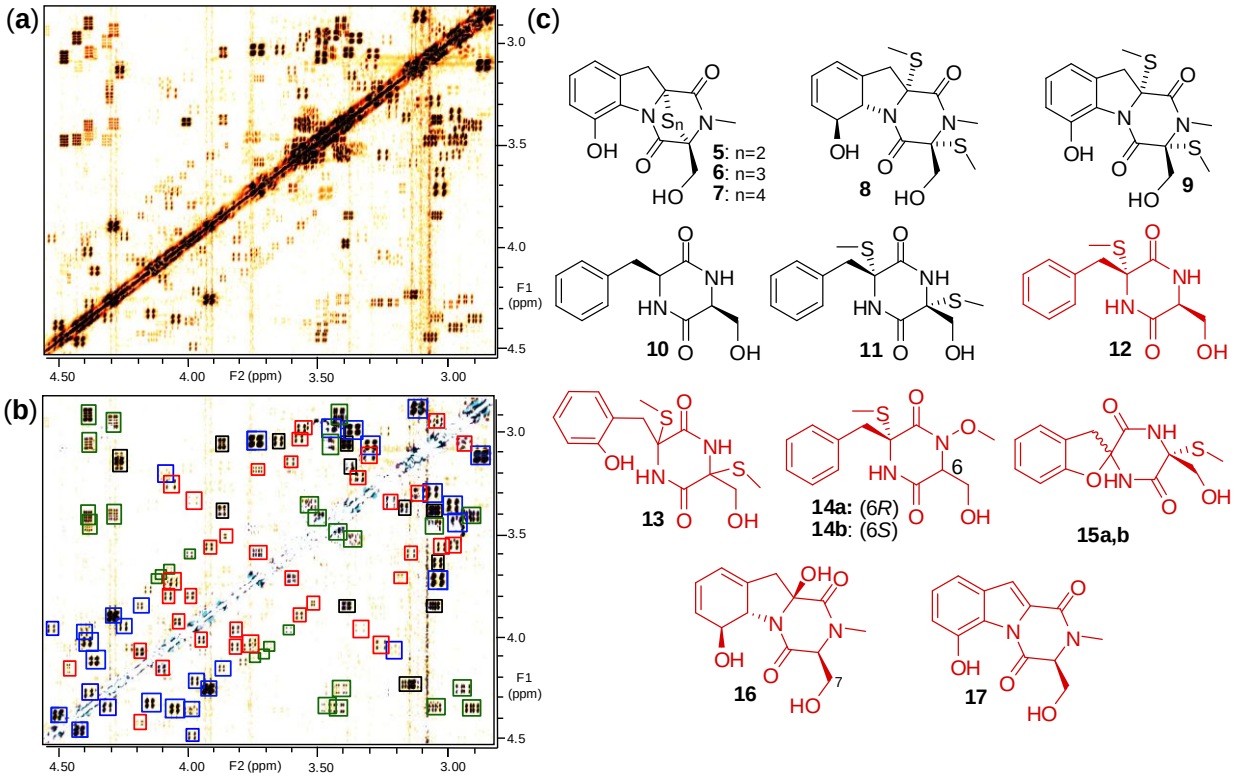


Figure 1.2: Identification of *gliZ*-dependent metabolites in pool 2. (**a**) DANS overlay of wild-type and Δ*gliZ* spectra, (**b**) Section of the wild-type dqfCOSY spectrum showing signals representing known gliotoxin derivatives (blue), the known **10** and **11** (black), and novel *gli*-dependent compounds (red). Several cross peaks representing metabolites present in both wild-type and Δ*gliZ* were not fully suppressed in the overlay due to chemical shift variation (green). (**c**) *gliZ*-dependent metabolites identified in wild-type pool 2. Structures shown in red represent novel metabolites. The relative configuration of **13**, which occurs as one single diastereomer, could not be determined, but likely corresponds to that of **11**.

Detailed analysis of *gliZ*-dependent signals in the wild-type pool 2 spectra revealed spin systems whose NMR data closely matched those of the known ETPs **1**-**3** (**Figure 1.1**) and the bis-methylsulfanyl derivative **8**,[24-26] in addition to a large number of unknown compounds (see **Figure 1.2** and Appendix A **Figure A.3** for DANS overlay spectra). For further structural assignments, the wild-type pool 2 sample was fractionated *via* semi-preparative HPLC and characterized fractions containing one or more of the *gliZ*-dependent compounds detected by DANS, using ^{1}H,^{13}C-HSQC, ^{1}H,^{13}C-

HMBC, and ROESY spectra as well as HPLC/ positive-ion electrospray ionization (ESI^+)-MS. These analyses confirmed the presence of the gliotoxins **1**, **3** and **8**, as well as their dehydro derivatives **5**, **7**, and **9**.[26-28] Cyclo(L-Phe-L-Ser) was also identified as a *gliZ*-dependent metabolite (**10**),[29] not previously reported from *A. fumigatus* extracts nor other gliotoxin-producing fungi, as well as bis-*N*-norgliovictin (**11**), a known *A. fumigatus* metabolite[30] not previously associated with the *gli* cluster.

All other *gliZ*-dependent metabolites detected in wild-type pool 2 represented novel compounds (**Figure 1.2**). A reoccurring structural motif in the *gliZ*-dependent metabolites of pool 2 is a cyclo(Phe-Ser) framework decorated with different oxidative functionality at the α-carbon positions. Amongst these functional groups, a methylsulfanyl bound *via* sulfur to either one or both of the α-carbons was frequently observed (compounds **12**-**15**). 3-bond $^{1}H,^{13}C$-HMBC correlations from the methylsulfanyl to corresponding α-carbons, along with additional $^{1}H,^{13}C$-HMBC correlations from the β methylene protons of either the Phe- or Ser-derived moieties helped to establish the diketopiperazines core structure.

Oxidation was not limited to the α-carbons. Compounds **14a/b** represent a pair of diastereomeric *N*-methoxy derivatives of **12**, as was demonstrated *via* $^{1}H,^{15}N$-HMBC showing *N*-methoxy protons coupling to *N*-1 and, in the case of compound **14a**, residual 1-bond *N*-4-H-4 coupling. ROESY through space correlations across the diketopiperazines ring established the relative stereochemistry of compounds **12**, **14a**/**b**, and **16**. Compounds **12**-**15** are accompanied by two indolopyrazines, **16** and **17**, representing nonsulfurized derivatives of gliotoxin. In addition, DANS revealed a second type of spirocyclic scaffold, the 6,9-diaza-1-oxaspiro-[4.5]decanes **15a** and **15b**. The two diastereomers **15a/b** appear to be derived from **13** *via* intramolecular substitution at C-3, as isolated samples of **13** slowly convert into mixtures of **15a** and **15b** (Appendix A, **Figure A.4**). HR-MS confirmed the molecular formulas of all new compounds (see Appendix A, **Table A.11** and **Table A.12** for spectroscopic data), and the *gliZ*-

dependence of their biosynthesis was corroborated *via* DANS and HPLC/ESI$^+$-MS in two independent replicates (Appendix A, **Figure A.5** and **Figure A.6**).

To test whether *gliZ* overexpression would reveal additional *gliZ*-dependent metabolites, the *A. fumigatus* wild-type metabolome was compared with that of a *gliZ*-overexpressing (OE) strain *via* DANS.[12] Although differences were observed in the relative amounts of some of the metabolites **1**-**17** in the OE strain, no new *gliZ*-dependent metabolites were detected. Next, the effect of deletion of two additional *gli* genes, *gliP* and *gliI* were investigated. DANS- and HPLC/ESI$^+$-MS-based comparison of *ΔgliP* and *ΔgliI* mutant metabolomes with wild-type showed that none of the *gliZ*-dependent compounds **1**-**17** are produced by these two mutant strains. These results further support that biosynthesis of **1**-**17** requires the *gli* cluster.

Gliotoxin Biosynthetic Pathway: Of the nineteen *gliZ*-dependent compounds identified, nine represent novel compounds, several of which feature structural motifs not previously associated with *gli* products. These structural features are of interest considering the putative pathway of gliotoxin biosynthesis in *A. fumigatus* (**Figure 1.3**).[5] Only the first and last steps in gliotoxin biosynthesis, the condensation between L-Phe and L-Ser by GliP and the oxidation of dithiol gliotoxin by GliT, have been elucidated,[6,17] and many aspects of the intervening steps remain unclear. In vitro experiments conducted with recombinant GliP demonstrated that GliP couples the amino acids L-Phe and L-Ser, producing a L-Phe-L-Ser-GliP intermediate.[6] Further observations indicated that GliP is capable of producing cyclo(L-Phe-L-Ser), but kinetic data suggested that the rate of cyclic dipeptide formation may be too low to be enzymatically relevant.[6] The comparative metabolomics work of this study have demonstrated that **10** is a major component of the *A. fumigatus* wild-type metabolome, with a gliotoxin-to-cyclo(L-Phe-L-Ser) molar ratio of roughly 2:1, and that this abundant production of **10** is *gliZ*-dependent. The abundance of **10** suggests that the *in vivo* rate of formation is much

greater than that *in vitro*, perhaps due to presence of additional factors, for example other *gli* components, *in vivo*. It has been proposed that **10** is an intermediate in gliotoxin biosynthesis;[7] however, the structural features of the *gliZ*-dependent metabolites identified are consistent with pathways involving tethered intermediates (**Figure 1.3**). Furthermore, the addition of synthetic **10** to *ΔgliP* cultures did not rescue the production of any of the *gliZ*-dependent metabolites as assessed by DANS and HPLC-MS (Appendix A, section 14), suggesting that **10** may not be a biosynthetic intermediate.

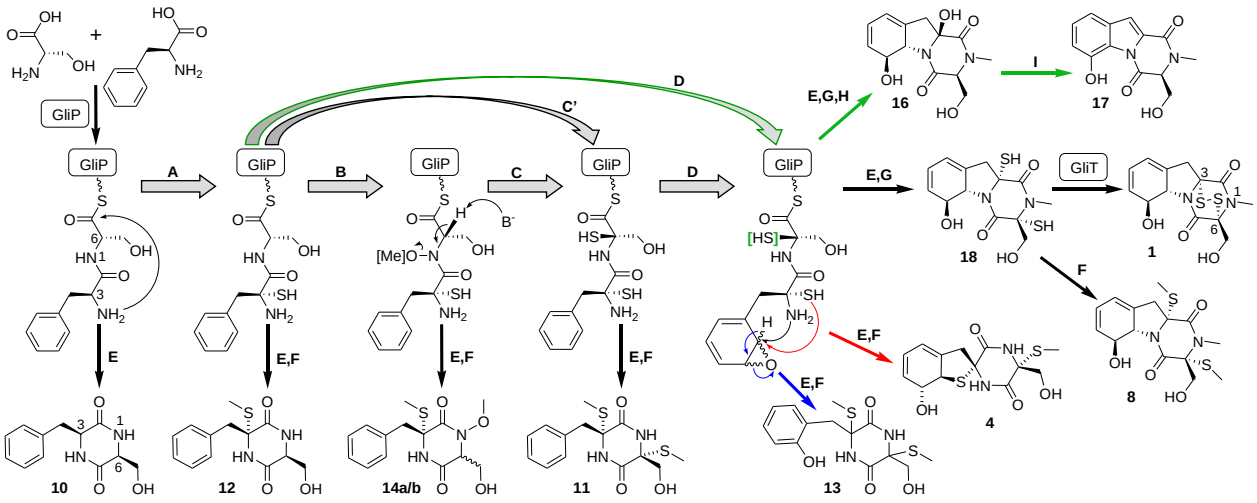


Figure 1.3: Proposed biogenesis of *gliZ*-dependent metabolites. Oxidation and sulfurization at C-3 (**A**) is followed by *N*-1-oxidation (**B**) and C-6 sulfurization (**C** or **C'**). Epoxidation (**D**) is followed by pyrollidine, thiophane, or phenol formation (black, red, and blue arrows, respectively). Aminolysis (**E**) of the resulting GliP-tethered intermediates and methylation of thiols (**F**) produce compounds **4**, **10**-**14a/b**. *N*-1-methylation (**G**) and aminolysis (**E**) result in formation of dithiol, which is *S*-methylated (**F**) to form **8** or oxidized by GliT[17] to produce gliotoxin, **1**. Hydrolysis (**H**) of species oxidized (and possibly sulfurized) only at C-3 (green arrows) produces **16**, which aromatizes (**I**) to form **17**.

All of the *gliZ*-dependent compounds **1**-**17** except **10** are oxidized at the α-carbon of Phe, and in the case of compound **12** only at that position, suggesting that oxidation

and sulfurization at C-3 occurs at an early stage in the biosynthesis following formation of the L-Phe-L-Ser dipeptide. Oxidation at C-3 could involve *gliC* or *gliF*, encoding putative cytochrome P450 monooxygenases. Sulfurization at C-3, yielding a C-3 thiol intermediate, likely involves GliG, recently demonstrated to possess glutathione *S*-transferase activity *in vitro*.[31] Compound **12** would then form *via* methylation of the thiol, either before or after aminolysis of the thioester. All sulfurized diketopiperazines identified in this study (except for the four ETPs) are *S*-methylated and thus likely derived from methylation of corresponding thiols, which is also suggested by the finding that, in *Gliocladium deliquescens*, gliotoxin is converted into bisdethiobis(methylthio)gliotoxin, **8**.[26]

The identification of compounds **11** and **14a/b** suggests that sulfurization at C-3 is followed by oxidation of the peptide backbone at C-6 and *N*-1, with *O*-methylation of the resulting hydroxamic acid derivatives by the putative *O*-methyltransferase GliM. Aside from spirocyclic **15a/b**, compounds **14a/b** are the only *gliZ*-dependent metabolites that occur as a pair of epimers, likely as a result of deprotonation/enolization at C-6. Oxidation at C-6, perhaps involving *gliC* or *gliF*, could occur independently from hydroxylation of *N*-1; however, sulfurization at C-6 could also be accomplished *via* dehydration of a hydroxamic acid intermediate followed by sulfurization of the resulting imine, as had been proposed previously on the basis of synthetic studies[32] and a shunt metabolite detected in a *ΔgliG* strain.[31] This mechanism is also suggested by the lack of *N*-hydroxylated or methoxylated compounds other than the C-6 non-sulfurized **14a/b**. Release and methylation of the bis-3,6-disulfurized dipeptide would then produce compound **11**.

As a next step, the disulfurized dipeptide appears to undergo epoxidation of the phenyl ring by one of two putative cytochrome P450 monooxygenases, GliC or GliF. The relative configurations observed for compound **4** strongly suggest that this oxidation step is not entirely stereoselective and produces both the (*13S*,*14R*)- and (*13R*,*14S*)-

diastereomers (**Figure 1.4**). Nucleophilic opening of the (*13*S,*14*R)-diastereomer by the C-3 amino group could lead to formation of the pyrrolidine ring of gliotoxin, whereas nucleophilic attack by the C-3 thiol results in formation of the thiophane ring in **4**. Alternatively, the phenyl ring is oxidized to form the *ortho*-substituted phenol found in compound **13**, perhaps *via* the shown epoxide intermediate. The biosyntheses of **13**, spirocyclic compound **4**, and the gliotoxins **1**-**3** and **8** are completed *via* diketopiperazine formation and *S*-methylation (in the case of **4** and **13**) or *N*-methylation, possibly involving *gliN*, and epidithio bridge formation via GliT in the case of the EPTs.[17] Lastly, compounds **16** and **17**, which both lack oxidation at C-6, appear to be derived from epoxidation of an intermediate oxidized only at C-3, followed by pyrrolidine and diketopiperazine formation. As the relative configuration at C-3 in **16** is opposite to that of C-3 in all identified C-3-sulfurized compounds, the C-3-OH group in **16** could be derived from substitution of a thiol or methylsulfanyl group. Elimination of water from **16** would form a dihydroindole that could easily aromatize to form **17**.

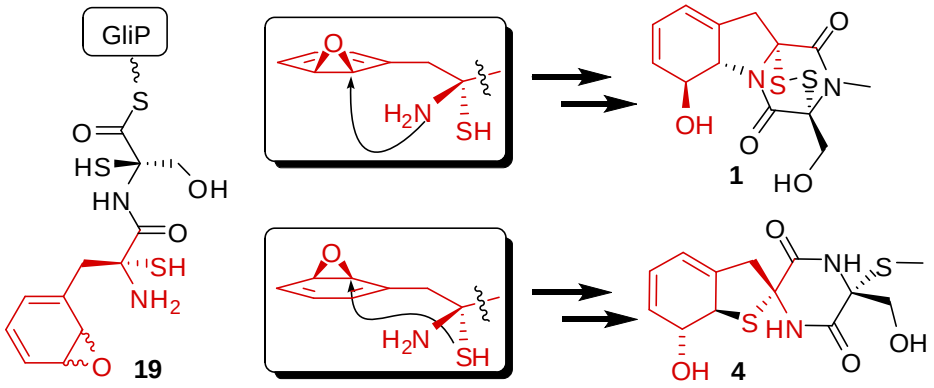


Figure 1.4: Proposed epoxide intermediate. Different configurations of putative epoxide intermediate **19** result in formation of either gliotoxin (**1**) or spirocyclic **4**.

Additional studies will be needed to test and further expand this model, to clarify the mechanism of sulfur incorporation,[31] and to determine whether tethered

intermediates or diketopiperazines are the substrates of the *gli*-encoded oxidases and methyltransferases. Furthermore, variation of growth conditions could induce biosynthesis of additional *gliZ*-dependent metabolites.

Conclusions: Comparison of the wild-type and *ΔgliZ* metabolomes *via* DANS revealed nine new compounds featuring several unexpected structural motifs, despite the fact that *A. fumigatus*'s metabolome, and specifically gliotoxin and its associated biosynthetic genes' role in virulence, had already been studied extensively. In particular, DANS facilitated detection of minor or unstable metabolites (e.g., **4** and **13**) missed by conventional analysis. Further investigations of the role of the *gli*-cluster in *A. fumigatus* virulence may benefit from the expanded knowledge of *gli*-associated structures and help clarify the biological roles of the newly identified compounds. Finally, these results suggest that the use of NMR-based comparative metabolomics for the examination of orphan PKS/NRPS gene clusters in microorganisms can significantly accelerate discovery of new structures and biosynthetic annotation.

Elaboration of the Gliotoxin Biosynthetic Pathway GliG, GliC, and GliI: After publication of the work highlighted above,[33] several recently studies have examined the details of sulfur incorporation in the gliotoxin biosynthetic pathway, including recombinant and mutant studies focusing on: GliG[31,34], GliC[34], and GliI[8] (**Figure 1.1**). The results of these studies corroborate some of the hypotheses developed in the DANS-based *gliZ* study pertaining to gliotoxin biosynthesis. **Figure 1.5** summarizes what has been reported about the small molecule transformations GliG, GliC and GliI can catalyze.

Two independent studies found that recombinant GliG can act as a glutathione *S*-transferase.[31,34] A Δ*gliG* mutant strain was shown to produce compound **20** (**Figure 1.5**), a highly oxidized cyclo(Phe-Ser) species containing: C-3 hydroxylation, *N*-4-

methoxy functionality, and a C-6 exocyclic methylene. Additionally in a study of the cytochrome P450-encoding genes *gliC* and *gliF*, **20** was found to be absent in a Δ*gliC* mutant strain and present in Δ*gliF* mutant, suggesting that GliC is responsible for the *N*- and/or α-carbon-oxidation and GliF may be implicated in oxidation of the Phe phenyl ring. The study went on to show that a synthetic compound (**21**), a cyclo(Phe-Ser) species hydroxylated at both α-carbons, undergoes glutathione conjugation catalyzed by GliG resulting in the *bis*-glutathione adduct **22**. Compound **22** was not observed in any reported mutant strain. An independent study probing the activity of GliG found that the recombinant enzyme was also able to transfer a glutathione moiety to 1,2-epoxy-3-(4-nitrophenoxy)-propane, 1-chloro-2,4-dinitrobenzene, and 3,4-dichloro-nitrobenzene.

A follow up study explored the role of the aminotransferase GliI in further processing following sulfur installation.[8] The Hertweck Laboratory explored the metabolites of a novel isolate of Δ*gliI* (independent of the Δ*gliI* strain described in Appendix A).[8] From a 140 L Δ*gliI* mutant culture 1.5 mg of a highly water soluble dicysteine analog of **22** (**23**) was isolated and fully characterized by NMR. Compound **23** was found to be converted to the corresponding dithiol (**24**) when exposed to recombinant GliI and coenzyme pyridoxal phosphate (PLP) *in vitro*. Additionally the authors show that this reaction results in the formation of pyruvic acid and ammonia, supporting the thiol formation mechanism outlined in **Figure 1.5**.

The GliG and GliI studies described above represent a firm step forward developing a more complete model of the gliotoxin biosynthetic pathway. The authors claim that the results of these mutant and *in vitro* experiments demonstrate that gliotoxin biosynthesis is largely carried out as controlled enzymatic modifications to untethered cyclo(Phe-Ser) analogs. However, one could argue that the results of these studies are also consistent with a tethered model and the observations of the DANS-based *gliZ* study. The metabolites isolated from the Δ*gliG* and Δ*gliI* mutants, **20** and **23** respectively, could have come from modification of a GliP-S-Ser-Phe-NH_2 intermediate,

followed by the off-loading mechanism proposed for the shunt metabolites depicted in **Figure 1.3**.

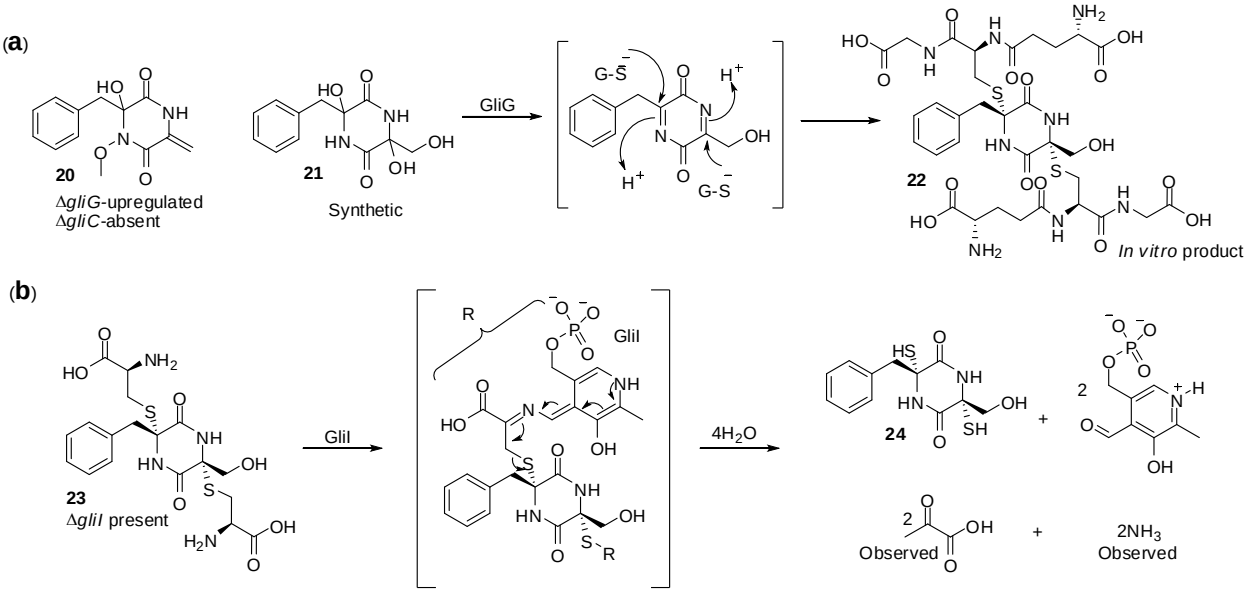


Figure 1.5: Overview of recently published work describing the chemistry of GliC, GliG and GliI. (**a**) Compound **20** was found in a Δ*gliG* mutant. Subsequently, compound **21** was tested as a substrate for GliG, demonstrating the addition of glutathione ($G\text{-}S^-$) affording **22**. (**b**) GliI was demonstrated to act as a carbon-sulfur lyase, converting **23** to **24** and yielding the side products pyruvic acid and ammonia.

Although it is clearly demonstrated that the recombinant enzymes can carry out transformations on the model diketopiperazines utilized in the GliG and GliI studies, no rate constants are offered for these reactions. As demonstrated by Balibar and Walsh, enzyme rate constants are important parameters to measure to provide evidence pertaining to the likelihood that a given enzymatic transformation occurs at a physiologically relevant rate.[6] This was demonstrated in their examination of GliP, found to produce cyclo(Phe-Ser) but not at a rate that would suggest that GliP functions to produce this chemical species. The importance of establishing enzyme reaction rates is echoed in the results of Davis *et al.*, described above, where GliG appears to have substrate flexibility, conjugating glutathione to several nitro-containing aromatic

substrates.[31] These findings suggesting that GliG could potentially also act on a the NRPS thioester-tethered dipeptide intermediate GliP-S-Ser-Phe-NH_2. As a potential complement to the *in vitro* work already done with recombinant GliG and GliI, further experiments incorporating GliP-S-Ser-Phe-NH_2 may provide additional insight into the tethered versus untethered gliotoxin biosynthetic models.

Examination a Wild-type and OE::rsmA A. fumigatus Mouse Infection Model: *Aspergillus fumigatus* uses several strategies to regulate secondary metabolic gene expression, including global (LaeA)[35] and gene-cluster specific (GliZ)[12] expression control. *A. fumigatus* poses a significant threat to immunocompromised hosts, and gliotoxin has received significant attention as a potential virulence factor involved in modulating *A. fumigatus's* efficacy of infection. Therefore, systematic studies of gliotoxin's biosynthetic regulatory pathways in *A. fumigatus* will help to deconvolute the relative importance of this metabolite in pathogenicity. The methodology and metabolites described in the DANS-*gliZ* study presents an analytical platform to monitor how the *gli* cluster, and overall gliotoxin biosynthesis, responds to perturbations in secondary metabolite gene regulations. Described herein is a study exploring one such perturbation, by the *gliZ* regulatory protein RsmA.

Recently a suppressor mutagenesis screen in *Aspergillus nidulans* identified *rsmA,* encoding a bZIP transcriptional enhancer, that when overexpressed could partially restore secondary metabolism in Δ*laeA*.[36] Overexpression of *rsmA* (*OE::rsmA*) increased sterigmatocystin production in *A. nidulans* more than 40-fold through transcriptional activation of *aflR*.[37] In continuation of the *gliZ* regulatory studies, an investigation of the *A. fumigatus rsmA* ortholog, also called *rsmA,* was undertaken both in culture and a mouse infection model.

Utilizing the established extraction protocols and HPLC/ESI^+-MS methodology established in the DANS-*gliZ* study described above,[33] compounds **1**-**17** were profiled in

cultures of wild-type and *OE::rsmA* (**Figure 1.6**). Standards derived from enriched and NMR-characterized samples of the *gliZ*-dependent metabolites (**1**-**17**) were utilized to confirm the identity of observed metabolites *via* HPLC retention time and MS fragmentation patterns. Comparison of peak areas for wild-type and *OE::rsmA* samples demonstrated that the majority of compounds (**1**, **3**, **10**, **12**, **14a/b**, **15a/b**, and **17**) are greater than 10-fold upregulated in *OE::rsmA* relative to wild-type. Compounds **2**, **8**, **9**, and **11** were 0.1-5-fold more abundant in *OE::rsmA* relative to wild-type.

In addition to culture analysis, the lungs of seven sacrificed female Swiss ICR mice, immunosuppressed *via* administration of cyclophosphamide and infected with wild-type, *OE::rsmA,* Δ*gliZ*, or *OE::rsmA A. fumigatus*, were profiled for the presence or absence of compounds **1**-**17**. Given the limited amount of sample (each pair of mouse lungs ~ cm^3), lungs from a given treatment set were pooled. To ensure the highest detection sensitivity for the metabolite profiling experiment, HPLC/ESI^+-SIMMS was carried out, restricting the mass spectral data acquisition to the most abundant ions for each metabolite profiled. Of compounds **1**-**17** only gliotoxin and cyclo(Phe-Ser) (**10**) were detected; however, compound **10** had not previously been described as a metabolite found in *A. fumigatus*-infected mice (**Figure 1.7**). The *OE::rsmA* sample contained about 2-fold more gliotoxin and **10** than wild-type. Survival experiments running in parallel to the lung metabolite profiling work showed that both wild-type- and *OE::rsmA*-infected mice do not survive as long as *OE::rsmA,* Δ*gliZ*-infected mice, and that there is no significant difference between the wild-type and *OE::rsmA*-infected mice survival times.

The greatly enhanced production of numerous *gli* cluster metabolites in *OE::rsmA* was striking and supported results from the DANS-*gliZ* study, showing that gliotoxin is not the only metabolite produced by the *gli* gene cluster. Two pertinent observations from this study include further support for GliT as an oxidase responsible for disulfide formation from dithiol-containing metabolites.[17] Notably, the lesser

accumulation of the methylsulfanyl containing compounds (**8**, **9**, and **11**) would be expected in the *OE::rsmA* strain, as *gliT* expression was particularly upregulated in this strain. GliT is also responsible for *A. fumigatus* self-resistance to gliotoxin, and likely other *gli* metabolites and, along with bis(methylthio)gliotoxin (**8**),[38] has been proposed as an antigen for diagnosis of invasive aspergillosis.[39] The profiling work here reveals an additional metabolite, cyclo(L-Phe-L-Ser) (**10**), which may also present an alternative molecule for invasive aspergillosis diagnosis. Compound **10**, along with gliotoxin, is produced in high amounts by the fungus both *in vivo* and *in vitro* and, furthermore, remains stable over time as would be required in a diagnostic setting.

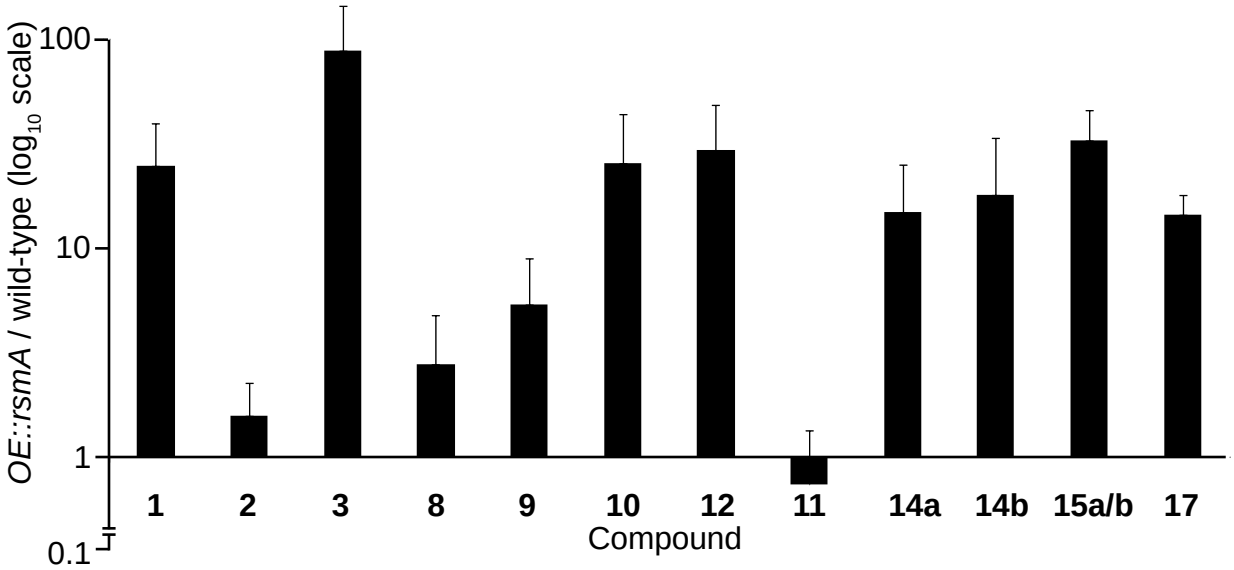


Figure 1.6: Profiling compounds 1-17 in *OE::rsmA* relative to wild-type. Relative amounts of compounds **1**-**17** in OE::*rsmA* compared to wild-type. Compounds **4**, **13**, and **16** were detected neither in OE::*rsmA* nor in wild-type cultures. Compounds **15a/b** and **17** were not detected in wild-type; values shown for these two compounds represent the measured S/N ratio as determined by their diagnostic mass ion chromatograms. Error bars represent one standard deviation (n = 3 cultures).

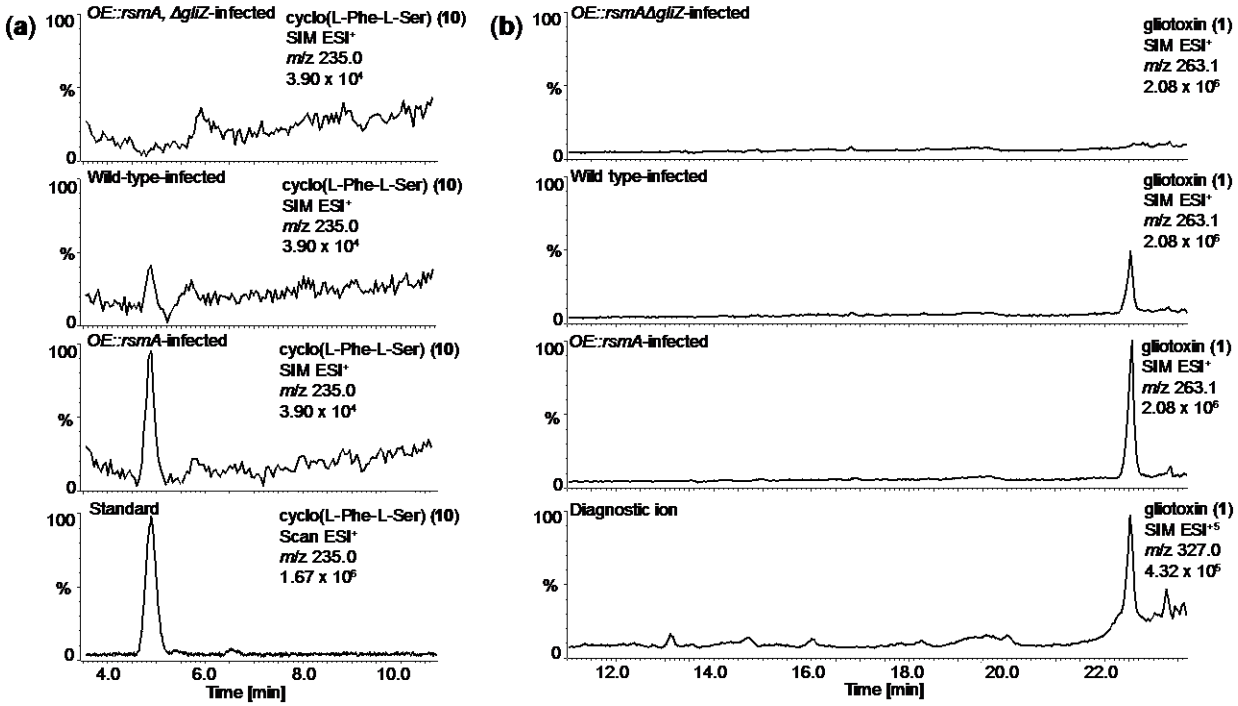


Figure 1.7: HPLC/ESI$^+$-SIMMS profiling results for *OE::rsmA,* Δ*gliZ*-, wild-type-, and *OE::rsmA*-infected mouse lung samples. (**a**) Compound **10** is present in *OE::rsmA*, somewhat less abundant in wild-type, and absent in *OE::rsmA,* Δ*gliZ.* (**b**) Gliotoxin is present in *OE::rsmA*, somewhat less in wild-type, and absent in *OE::rsmA,* Δ*gliZ*.

REFERENCES

(1) Nierman, W. C. *et al. Nature* **2005**, *438*, 1151.

(2) Rohlfs, M.; Albert, M.; Keller, N. P.; Kempken, F. *Biol Lett* **2007**, *3*, 523.

(3) Kupfahl, C.; Michalka, A.; Lass-Florl, C.; Fischer, G.; Haase, G.; Ruppert, T.; Geginat, G.; Hof, H. *Int J Med Microbiol* **2008**, *298*, 319.

(4) Lewis, R. E.; Wiederhold, N. P.; Chi, J.; Han, X. Y.; Komanduri, K. V.; Kontoyiannis, D. P.; Prince, R. A. *Infect Immun* **2005**, *73*, 635.

(5) Gardiner, D. M.; Howlett, B. J. *FEMS Microbiol Lett* **2005**, *248*, 241.

(6) Balibar, C. J.; Walsh, C. T. *Biochemistry* **2006**, *45*, 15029.

(7) Fox, E. M.; Howlett, B. J. *Mycol Res* **2008**, *112*, 162.

(8) Scharf, D. H.; Chankhamjon, P.; Scherlach, K.; Heinekamp, T.; Roth, M.; Brakhage, A. A.; Hertweck, C. *Angew Chem Int Ed Engl* **2012**, *51*, 10064.

(9) Cramer, R. A., Jr.; Gamcsik, M. P.; Brooking, R. M.; Najvar, L. K.; Kirkpatrick, W. R.; Patterson, T. F.; Balibar, C. J.; Graybill, J. R.; Perfect, J. R.; Abraham, S. N.; Steinbach, W. J. *Eukaryot Cell* **2006**, *5*, 972.

(10) Kupfahl, C.; Heinekamp, T.; Geginat, G.; Ruppert, T.; Hartl, A.; Hof, H.; Brakhage, A. A. *Mol Microbiol* **2006**, *62*, 292.

(11) Sugui, J. A.; Pardo, J.; Chang, Y. C.; Zarember, K. A.; Nardone, G.; Galvez, E. M.; Mullbacher, A.; Gallin, J. I.; Simon, M. M.; Kwon-Chung, K. J. *Eukaryot Cell* **2007**, *6*, 1562.

(12) Bok, J. W.; Chung, D.; Balajee, S. A.; Marr, K. A.; Andes, D.; Nielsen, K. F.; Frisvad, J. C.; Kirby, K. A.; Keller, N. P. *Infect Immun* **2006**, *74*, 6761.

(13) Challis, G. L. *J Med Chem* **2008**, *51*, 2618.

(14) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Biol* **2010**, *15*, 38.

(15) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straightt, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *PNAS USA* **2007**, *104*, 1506.

(16) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *PNAS USA* **2009**, *106*, 7708.

(17) Scharf, D. H.; Remme, N.; Heinekamp, T.; Hortschansky, P.; Brakhage, A. A.; Hertweck, C. *J Am Chem Soc* **2010**, *132*, 10136.

(18) Schrettl, M.; Carberry, S.; Kavanagh, K.; Haas, H.; Jones, G. W.; O'Brien, J.; Nolan, A.; Stephens, J.; Fenelon, O.; Doyle, S. *PLoS Pathog* **2010**, *6*, e1000952.

(19) Zhang, F.; Bruschweiler, R. *Angew Chem Int Ed Engl* **2007**, *46*, 2639.

(20) Fan, T. W.; Yuan, P.; Lane, A. N.; Higashi, R. M.; Wang, Y.; Hamidi, A. B.; Zhou, R.; Guitart, X.; Chen, G.; Manji, H. K.; Kaddurah-Daouk, R. *Metabolomics* **2010**, *6*, 165.

(21) Claridge, T. D. W. *High-resolution NMR techniques in organic chemistry*; 2nd ed.; Elsevier: Amsterdam ; Boston, **2009**.

(22) Taggi, A. E.; Meinwald, J.; Schroeder, F. C. *J Am Chem Soc* **2004**, *126*, 10364.

(23) Gronquist, M.; Meinwald, J.; Eisner, T.; Schroeder, F. C. *J Am Chem Soc* **2005**, *127*, 10810.

(24) Fukuyama, T.; Nakatsuka, S.; Kishi, Y. *Tetrahedron* **1981**, *37*, 2045.

(25) Kirby, G. W.; Rao, G. V.; Robins, D. J. *J Chem Soc Perk T 1* **1988**, 301.

(26) Kirby, G. W.; Robins, D. J.; Sefton, M. A.; Talekar, R. R. *J Chem Soc Perk T 1* **1980**, 119.

(27) Safe, S.; Taylor, A. *J Chem Soc C* **1970**, 432.

(28) Hanson, J. R.; Oleary, M. A. *J Chem Soc Perk T 1* **1981**, 218.

(29) Campo, V. L.; Martins, M. B.; da Silva, C. H. T. P.; Carvalho, I. *Tetrahedron* **2009**, *65*, 5343.

(30) Zhao, W. Y.; Zhu, T. J.; Han, X. X.; Fan, G. T.; Liu, H. B.; Zhu, W. M.; Gu, Q. Q. *Nat Prod Res* **2009**, *23*, 203.

(31) Davis, C.; Carberry, S.; Schrettl, M.; Singh, I.; Stephens, J. C.; Barry, S. M.; Kavanagh, K.; Challis, G. L.; Brougham, D.; Doyle, S. *Chem Biol* **2011**, *18*, 542.

(32) Herscheid, J. D. M.; Nivard, R. J. F.; Tijhuis, M. W.; Ottenheijm, H. C. J. *Journal of Organic Chemistry* **1980**, *45*, 1885.

(33) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(34) Scharf, D. H.; Remme, N.; Habel, A.; Chankhamjon, P.; Scherlach, K.; Heinekamp, T.; Hortschansky, P.; Brakhage, A. A.; Hertweck, C. *J Am Chem Soc* **2011**, *133*, 12322.

(35) Bok, J. W.; Keller, N. P. *Eukaryot Cell* **2004**, *3*, 527.

(36) Shaaban, M. I.; Bok, J. W.; Lauer, C.; Keller, N. P. *Eukaryot Cell* **2010**, *9*, 1816.

(37) Yin, W. B.; Amaike, S.; Wohlbach, D. J.; Gasch, A. P.; Chiang, Y. M.; Wang, C. C.; Bok, J. W.; Rohlfs, M.; Keller, N. P. *Mol Microbiol* **2012**, *83*, 1024.

(38) Domingo, M. P.; Colmenarejo, C.; Martinez-Lostao, L.; Mullbacher, A.; Jarne, C.; Revillo, M. J.; Delgado, P.; Roc, L.; Meis, J. F.; Rezusta, A.; Pardo, J.; Galvez, E. M. *Diagn Micobiol Infect Dis* **2012**, *73*, 57.

(39) Shi, L. N.; Li, F. Q.; Huang, M.; Lu, J. F.; Kong, X. X.; Wang, S. Q.; Shao, H. F. *BMC Microbiol* **2012**, *12*, 11.

CHAPTER 2

A BACKUP PLAN FOR SELF-PROTECTION: S-METHYLATION OF HOLOMYCIN BIOSYNTHETIC INTERMEDIATES IN STREPTOMYCES CLAVULIGERUS

Abstract: Biosynthesis of the dithiolopyrrolone antibiotic holomycin in *Streptomyces clavuligerus* involves the closure of a pair of enethiols to a cyclic disulfide. It has been demonstrated that the dithiol oxidase HlmI is responsible for the disulfide formation and this enzyme also plays a role in self-protection. This study examines how *S. clavuligerus* deals with the proposed toxic dithiol intermediates when *hlmI* is deleted. Comparative metabolomics employing differential NMR spectroscopy and mass spectrometry were used to profile the metabolomes of *hlmI* deletion mutants along with the wild-type strain and a holomycin-overproducing strain. A number of metabolites unique to the Δ*hlmI* strains were identified. In these metabolites the enethiols have been incapacitated by a combination of mono- and di-*S*-methylation. Also observed was an intriguing dimeric thioether adduct in low quantities in the wild-type strain and at much higher levels in the Δ*hlmI* strains. The structures of these novel metabolites highlight the reactivity of the dihydrodithiolopyrrolone scaffold. Furthermore, bioassays suggest that modification of the enethiol by *S*-alkylation provides a host strategy for detoxification, one that is shared amongst multiple species producing such bioactive disulfide natural products.

Introduction: Holomycin is a member of the *N*-acylated dithiolopyrrolone class of natural product antibiotics reported to interfere with bacterial RNA metabolism.[1-3] The antibiotic scaffold contains an unusual bicyclic core (**1**, **Figure 2.1**) comprising a cyclic ene-disulfide fused to an aminopyrrolone ring. While a much more compact framework than the mycotoxin gliotoxin or the FDA approved histone deacetylase inhibitor FK_{228}

(**Figure 2.1**), holomycin joins these molecules in that it possesses a disulfide-bridged heterocycle which is essential to its biologic activity. Disulfides in these molecules are likely to be precursors, reduced in cells to their active dithiol forms.[4,5] In *Streptomyces clavuligerus*, holomycin biosynthesis proceeds *via* the formation of a tethered Cys-Cys-*S*-enzyme intermediate on a Non-ribosomal peptide synthetase (NRPS) template. Subsequent oxidation by four flavoproteins[6] likely contributes to the formation of a pair of enethiols (one endocyclic, the other exocyclic) that undergo enzymatic oxidation to close the disulfide imbedded in the bicyclic scaffold. For gliotoxin it has been established that the enzyme GliT, which acts late to convert dithio-gliotoxin (**Figure 2.1**, **Figure 1.3**) to its isolated disulfide, plays a self-protective role in the producing Aspergilli.[7] *In Aspergillus fumigatus* the deletion of *gliT* resulted in mutants that are much more sensitive to gliotoxin when compared to the wild-type. This heightened sensitivity is presumably due to decreased ability of the producer to sequester gliotoxin as the disulfide, resulting in consequent buildup of the active dithiol form. Toxicity of the dithiol form of gliotoxin may result from modification of host proteins or redox-cycling, to generate reactive oxygen species.[8] By analogy, it was recently shown that *hlmI*, part of the holomycin (*hlm*) biosynthetic gene cluster in *S. clavuligerus*, encodes a comparable flavin-dependent dihydroholomycin (**2**, **Figure 2.1**) oxidase that converts **2** to the bicyclic oxidized form (**1**).[9] To examine the *hlm* pathway in detail one can take advantage of the dramatic upregulation of the *hlm* pathway in the *S. clavuligerus* ORF15 deletion mutant (ΔORF15). ORF15 is involved in the biosynthesis of another antibiotic, clavulanate, but the ΔORF15 strain overproduces holomycin ten- to 100-fold relative to the wild-type.[10] In both *S. clavuligerus* wild-type and ΔORF15 background, deletion of the *hlmI* gene resulted in a 100- to 1000-fold decrease in production of holomycin and a significantly heightened sensitivity to exogenously added holomycin.[9] Unresolved was the fate of any presumptively toxic dihydroholomycin accumulating in the Δ*hlmI* and Δ*hlmI*/Δ*ORF15* strains and how these mutants might deal with such a

dithiol load. Herein, LC/ESI$^+$-MS- and NMR-based comparative metabolomics strategies are described for the identification of dihydroholomycin and dihydroholothin derivatives that accumulate in the Δ*hlmI* and Δ*hlmI*/ΔORF15 strains and methodology to investigate the related bacterial self-protection mechanisms.

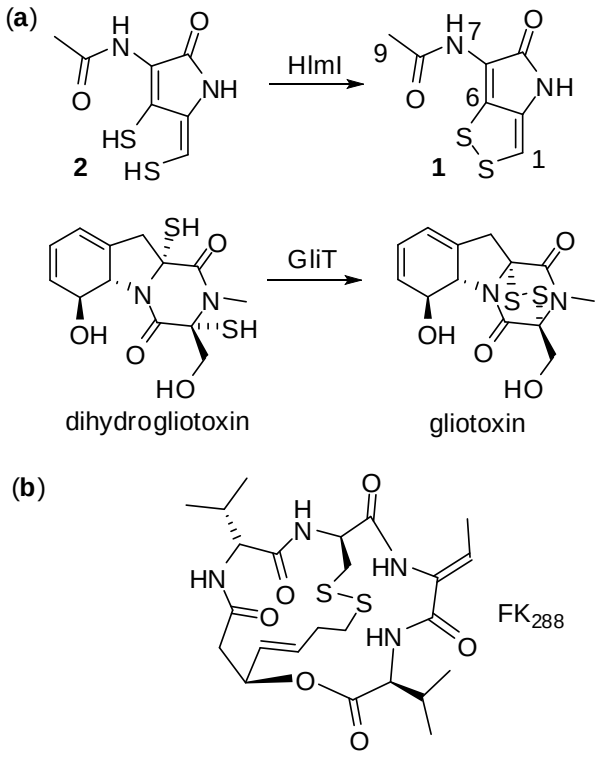


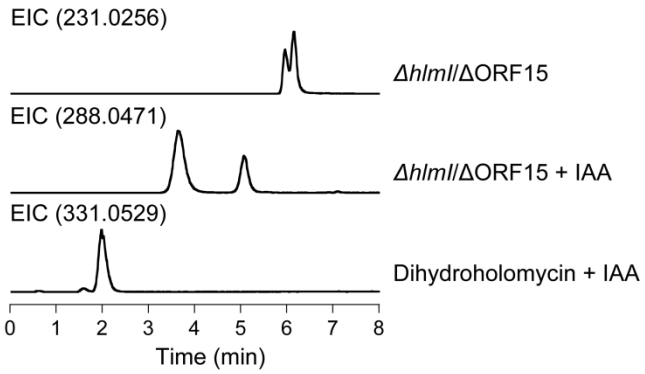
Figure 2.1: Compounds containing biologically active disulfides. (**a**) The conversion of the dihydro-forms of antibiotics holomycin (**1**) and gliotoxin to their disulfide forms by dithiol oxidases HlmI and GliT, respectively. (**b**) The structure of the drug FK228, the biological activity of which also depends on an intramolecular disulfide.

Detection of S-Methylated Forms of Dihydroholomycin: To gain insight into whether dihydroholomycin intermediates build up in the *hlmI* deletion mutants, the metabolomes of the Δ*hlmI*/ΔORF15 and ΔORF15 strains were compared through HPLC/ESI$^+$-high resolution (HR) MS. Peaks with masses corresponding to dihydroholomycin (*m*/z calcd for $C_7H_9N_2O_2S_2$: 217.0100) and dihydroholothin (*m*/z calcd for $C_5H_7N_2OS_2$: 174.9994) were not observed. Instead, two strong peaks (*m*/z found: 231.0244 $[M+H]^+$, *m*/z calcd

for $C_8H_{11}N_2O_2S_2$: 231.0256) were present in Δ*hlmI*/ΔORF15 but not ΔORF15 (**Figure 2.2** top trace). The mass difference of 14.0157 amu relative to dihydroholomycin indicated the addition of a CH_2 unit, suggesting possible *S*-methylation (**3**, **Figure 2.2** regio and/or cis/trans stereoisomers) or addition of a methyl group to the *N*-3 amide nitrogen of the dithiolopyrrolone scaffold, as observed in the structure of the related natural product thiolutin.[11] As a first pass to distinguish between *S*- or *N*-methylation, an excess of thiol-alkylating agent iodoacetamide (IAA) was added to the culture supernatant of Δ*hlmI*/ΔORF15. After incubation at room temperature for one hour, the *m*/z 231.0256 peak disappeared and two well-separated new peaks emerged, both of which had molecular ions (*m*/z found: 288.0466 $[M+H]^+$, *m*/z calcd for $C_9H_{11}N_3O_3S_2$: 288.0471) suggestive of two IAA adducts (**Figure 2.2**, middle trace). Additionally, both new peaks retained the signature UV absorption at 370 nm, suggesting that they are related to, and retain the chromophore of holomycin[12] (dihydroholomycin has a unique UV absorption at 330 nm, different from that of holomycin or alkylated dihydroholomycin, see **Figure B.2**, Appendix B). However, no bis-IAA adduct was detected, whereas a control reaction with authentic dihydroholomycin led exclusively to formation of the bis-IAA adduct (**Figure 2.2**, bottom trace, *m*/z found: 331.0553 $[M+H]^+$, *m*/z calcd for $C_{11}H_{15}N_4O_4S_2$: 331.0529). The proposed structure for the bis-IAA adduct (**4**) is shown in **Figure 2.2**. These findings indicate that Δ*hlmI*/ΔORF15 does not yield detectable quantities of dihydroholomycin and that at least two mono-*S*-methylated dihydroholomycin isomers (**Figure 2.2**, top trace) are present in its stead.

Comparative Metabolome Analysis through NMR Spectroscopy: To further characterize biosynthetic intermediates or shunt metabolites that accumulate as a result of the *hlmI* deletion, NMR spectroscopy-based comparison of the Δ*hlmI*/ΔORF15 and ΔORF15 metabolomes was employed. Differential analyses of 1D and 2D NMR spectra (DANS) enables largely unbiased comparisons of related metabolome samples to be

(**a**)



(**b**)

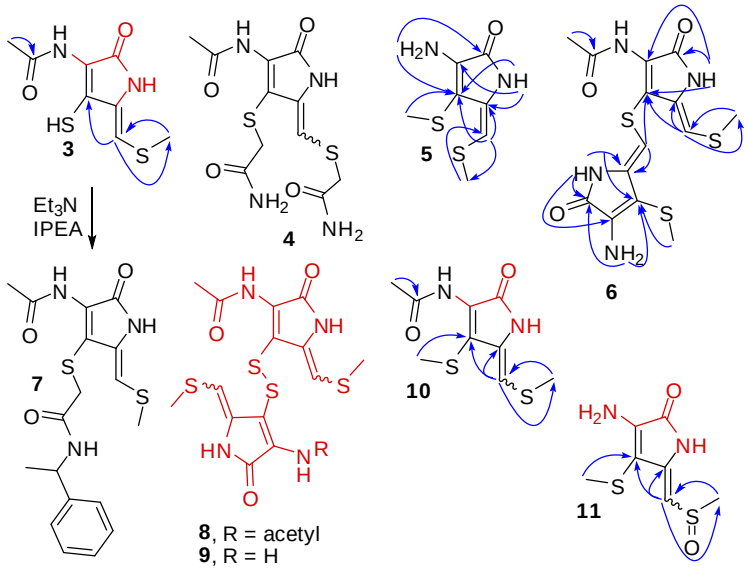


Figure 2.2: Identification and characterization of metabolites related to the *hlm* pathway in the Δ*hlmI*/ΔORF15 mutant. (**a**) Characterization of the holomycin + 14.0157 amu metabolite (*m*/z of 231.0256) in Δ*hlmI*/ΔORF15 by HPLC/ESI$^+$-MS and chemical derivatization with iodoacetamide (IAA). Shown are the extracted ion chromatograms (EIC) of Δ*hlmI*/ΔORF15 (top, $[M+H]^+$, 231.0256), Δ*hlmI*/ΔORF15 treated with IAA (middle, $[M+H]^+$, 288.0471), and dihydroholomycin standard treated with IAA (bottom, $[M+H]^+$, 331.0529). (**b**) The structures of *hlm* biosynthetic pathway intermediates, semi-synthetic derivative, and select disulfide containing metabolites. Compounds **3**, **5**, **6**, and **8**-**11** are all novel compounds upregulated in the Δ*hlmI*/ΔORF15 mutant. Compounds, or portions of compounds, colored in red represent structural elements that were supported by UPLC/ESI$^+$-HRMS(MS) analysis, but could not fully be characterized by NMR spectroscopy. Blue arrows represent HMBC correlations that support the structural assignments in these compounds. IPEA: 2-iodo-*N*-(1-phenylethyl)acetamide.

made, and has previously been employed in microbial[13,14] and animal systems[15] to identify metabolites associated with genetic or environmental perturbations.[16] Initial ^{1}H spectra acquired for crude extracts of Δ*hlmI*/ΔORF15 and ΔORF15 revealed at least nine Δ*hlmI*/ΔORF15-unique singlets between 5.5 and 6.7 ppm (**Figure 2.3**, for full spectra see **Figures B.5** and **B.6**). DANS comparison of dqfCOSY spectra acquired for extracts of Δ*hlmI*/ΔORF15 and ΔORF15 revealed no additional differential proton signals. To identify the Δ*hlmI*/ΔORF15-upregulated metabolites through ^{1}H,^{13}C -HMQC and ^{1}H,^{13}C-HMBC spectra, partial fractionation of the Δ*hlmI*/ΔORF15 and ΔORF15 samples was required as the compounds of interest accounted for less than 2% of the total Δ*hlmI*/ΔORF15 metabolome (extracted from 500 mL of a two-day culture at an $OD_{600\ nm}$ value of ~16). Reversed-phase flash chromatography of the Δ*hlmI*/ΔORF15 and ΔORF15 extracts yielded three metabolite pools, of which pool 1 contained the most polar metabolites and pool 3 the least polar metabolites. ^{1}H spectra were acquired for these six pools (Appendix B, section 3).

Identification of a Methylated Asymmetric Dimeric Holomycin Derivative: Comparison of ^{1}H spectra corresponding to pool 3 Δ*hlmI*/ΔORF15 and pool 3 ΔORF15 samples showed two groups of singlets (three peaks between 6.3-5.7 ppm and four peaks between 2.5-2.0 ppm) dramatically upregulated in the Δ*hlmI*/ΔORF15 spectrum (**Figure B.11**). ^{1}H,^{13}C-HMBC and ^{1}H,^{13}C-HMQC spectra acquired for Δ*hlmI*/ΔORF15 pool 3 revealed that the Δ*hlmI*/ΔORF15-upregulated singlets belong to two new compounds with correlation patterns and chemical shifts similar to those observed for holomycin (**1**) and dihydroholomycin (**2**). Based on ^{1}H,^{13}C-HMBC and an additional ROESY spectrum, one of the two unknown compounds was identified as bis-*S*-methylated dihydroholothin (**5**) (**Figure 2.2**). The remaining Δ*hlmI*/ΔORF15-dependent analyte in pool 3, like compound **5**, appeared to contain a primary, non-acylated amine substituent that showed 3-bond ^{1}H,^{13}C-HMBC correlations from its NH_2-protons to an.

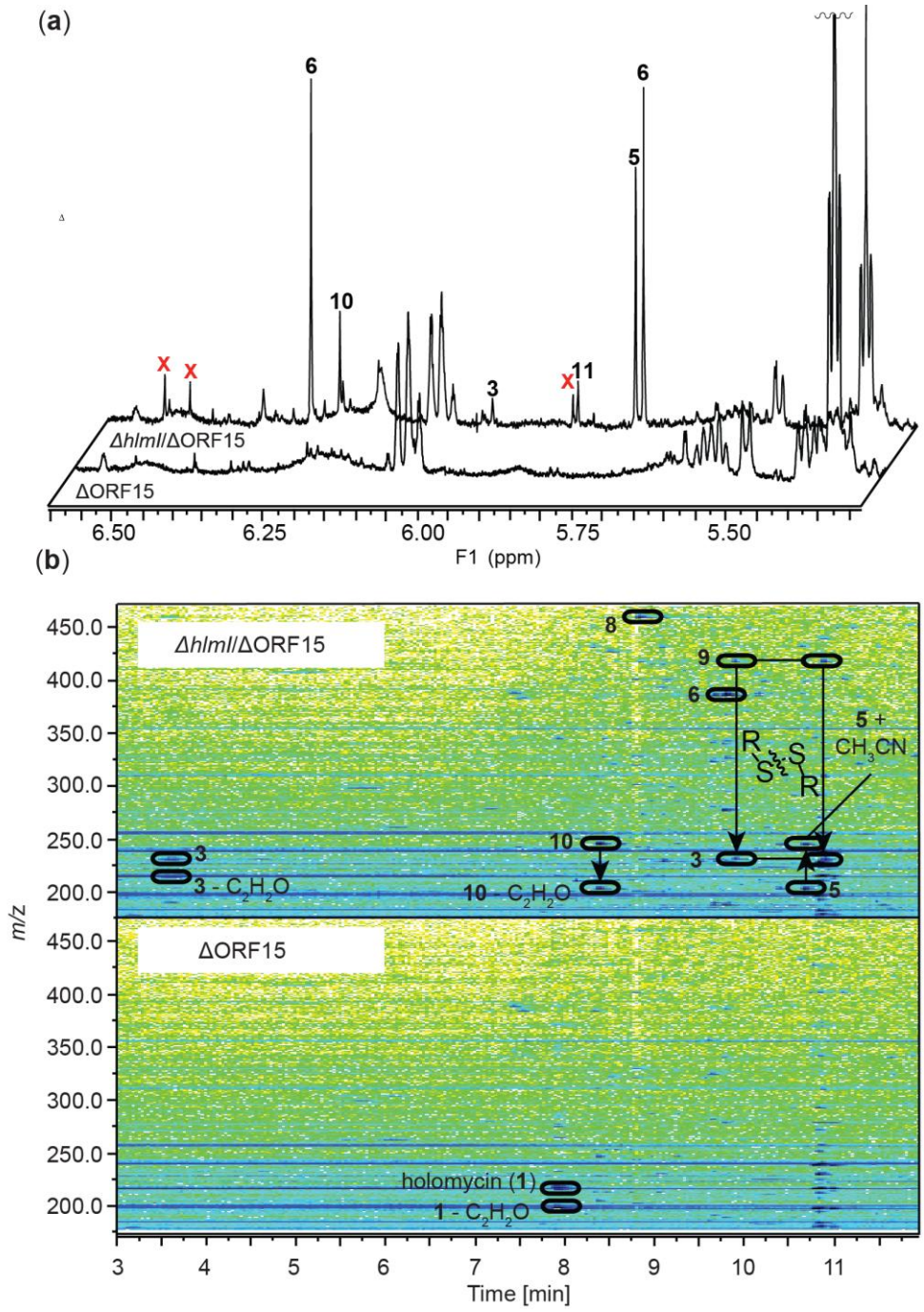


Figure 2.3: Comparative ^{1}H NMR and 2D HPLC/ESI^{+}-MS analysis of crude extracts of Δ*hlmI*/ΔORF15 and ΔORF15. (**a**) Region of ^{1}H spectra containing proton singlets diagnostic for H-1 or H-1' of compounds **3**, **5**, **6**, and **8**-**11**. Peaks marked with a red X remain uncharacterized, but may belong to compounds **8** and/or **9**. (**b**) Comparison of mass spectra (*m*/z 175-470) versus retention time (3-12 min) plots corresponding to extracts of the two mutants Δ*hlmI*/ΔORF15 and ΔORF15 revealed a number of cross peaks diagnostic for compounds **3**, **5**, **6**, and **8**-**11** that are exclusively found, or dramatically upregulated, in Δ*hlmI*/ΔORF15.

amide carbonyl carbon and a methylsulfanyl-substituted olefin carbon. Notably, an exocyclic olefinic proton, analogous to H-1 of **5**, coupled with carbons of the enamine ring system and a second *N*-7-acetylated pyrrolone ring system suggesting a thioether linkage. ROESY correlations between H-1 and H-1' further supported these structural assignments and determined the stereochemistry of this compound (**6**, **Figure 2.2**). Subsequent acquisition of UPLC/ESI$^+$-HRMS data confirmed the molecular formulae for both **5** and **6**. The pool 3 ^{1}H,^{13}C-HMBC spectrum further indicated the presence of at least four minor stereoisomers of compound **6**.

S-Methylated Dihydroholomycin Derivatives: Comparison of the pool 1 ^{1}H spectra revealed three ^{1}H singlets (5.97, 5.90, and ~2.3 ppm) in the Δ*hlmI*/ΔORF15 spectrum that were absent from the ΔORF15 spectrum. ^{1}H,^{13}C-gHMBC spectra indicated that the two protons at 5.97 and 5.90 ppm belong to similar ^{1}H,^{13}C spin systems. Both protons couple with carbons belonging to a methylsulfanyl moiety (~17 ppm) and appear to be attached to carbons that are part of a holomycin-like olefinic system. The proton at 5.97 ppm showed an additional correlation to another olefinic carbon that does not couple with any other protons. Additionally, the pool 1 Δ*hlmI*/ΔORF15 ^{1}H,^{13}C-gHMBC spectrum showed two major correlations from methyl protons (~1.8 ppm) to carbons at ~170 ppm, suggesting the presence of acetylated enamines. The absence of any additional methylsulfanyl correlations in the ^{1}H,^{13}C-gHMBC spectrum further supported the hypothesis that these compounds are a pair of monomethyl dihydroholomycin stereoisomers (**3**) and possibly its trans-stereoisomer, (**Figure 2.2**). Because of poor spectroscopic line shapes of the proton signals in the ^{1}H,^{13}C-HMBC spectrum, not all carbons in the proposed structures **3** could be detected. Derivatization with 2-iodo-*N*-(1-phenylethyl)acetamide (IPEA) under basic conditions led to isolation of a compound whose ^{1}H, ^{1}H,^{13}C-HMQC and ^{1}H,^{13}C-HMBC and UPLC/ESI$^+$-HRMSMS spectra were consistent with the structure **7**, confirming the presence of **3** in the Δ*hlmI*/ΔORF15 pool

1 sample. The trans-stereoisomer of **7** was not observed. UPLC/ESI$^+$-HRMS analysis of pool 1 and the IPEA derivative further supported these structural assignments. In addition, high- and unit-resolution HPLC/ESI$^+$-MS analysis of Δ*hlmI*/ΔORF15 extracts showed several chromatographic peaks whose mass spectra suggested that they represent disulfide-bridged dimers of **3** such as **8** and **9**, which likely formed as a result of nonenzymatic oxidation during processing of the samples (Appendix B, section 8).

Pool 2 Comparative Metabolomics: Comparison of ^{1}H spectra acquired for the pool 2 Δ*hlmI*/ΔORF15 and ΔORF15 samples (**Figure B.9** and **B.10**) showed several singlets (^{1}H, 6.22, 5.83, 5.50, and six between 2.89-1.97 ppm) upregulated in Δ*hlmI*/ΔORF15 relative to ΔORF15. A ^{1}H,^{13}C-gHMBC spectrum acquired for the pool 2 Δ*hlmI*/ΔORF15 sample reveled two major spin systems belonging to the pool 2 Δ*hlmI*/ΔORF15-upregulated compounds. The most downfield signal (^{1}H, 6.22 ppm) showed ^{1}H-^{13}C correlations and chemical shifts similar to those observed for **5**; where an exocyclic olefinic proton couples with a ^{13}C of a methylsulfanyl (16.6 ppm) and two endocyclic olefinic ^{13}C atoms; one weak correlation (^{13}C, 132.3 ppm, two bond) and one strong correlation (^{13}C, 133.1, three bond). Additionally, three methyl protons coupled to, and only to, a ^{13}C atom at 133.1 ppm. These initial data suggested a bis-methylated dihydroholomycin structure (**10**). However, the ^{1}H-^{13}C correlations needed to confirm this structural hypothesis were not observed. The second Δ*hlmI*/ΔORF15-upregulated spin system of pool 2 had markedly different ^{13}C chemical shifts, but showed a similar ^{1}H-^{13}C coupling pattern as the first spin system. ^{1}H,^{13}C-HMBC spectra acquired for this metabolite showed a proton at 5.83 ppm that coupled to a ^{13}C atoms at: 145.2, 95.7, and 40.9 ppm. Intriguingly, the 40.9 ppm ^{13}C atom belonged to a methyl group (^{1}H, 2.47 ppm) that coupled exclusively with the carbon bearing the 5.83 ppm proton (^{13}C, 111.5 ppm). Three methyl protons (2.16 ppm) also coupling to, and only to, the ^{13}C atom at 95.7 ppm. Taken together, the second Δ*hlmI*/ΔORF15-upregulated compound of pool 2

appeared to represent a structure similar to **10** (**Figure 2.2**), where the sulfur of the C-1 methylsulfanyl had been oxidized to a methyl sulfoxide substituent (**11**). As with compound **10**, not all ^{1}H-^{13}C correlations needed to fully characterize this compound were observed.

To reduce the dynamic range of Δ*hlmI*/ΔORF15 pool 2, **10** and **11** were enriched by flash chromatography (Appendix B, section 5) and ^{1}H spectra were acquired for select fractions (Appendix B, section 5). Although samples enriched with **10** and **11** were obtained, as was the case with compound **3**, ^{1}H line shape had worsened for these samples relative to the pre-fractionated pool 2 sample. Therefore, to test the hypothesized structures of **10** and **11** UPLC/ESI-HRMS studies were undertaking to further probe these metabolites, providing data consistent with the proposed structures of **10** and **11** (*m*/z found: 245.0401 $[M+H]^+$, *m*/z calcd for $C_9H_{13}N_2O_2S_2$: 245.0413 for **10**, and *m*/z found: 219.0240 $[M+H]^+$, *m*/z calcd for $C_7H_{11}N_2O_2S_2$: 219.0256 for **11**).

Targeted Metabolite Profiling of Wild-Type, ΔORF15, and Δhlml S. clavuligerus Strains: Having established several chemical species unregulated in Δ*hlmI*/ΔORF15 relative to ΔORF15, theses metabolites were profiled for their presence/absence in wild-type strain, and their relative quantities were assessed. For this purpose, targeted HPLC/ESI^+-MS-based profiling was utilized, searching for the diagnostic mass spectral fragmentation patterns and retention times established for the Δ*hlmI*/ΔORF15-dependent metabolites **3**, **5**, **6**, and **8**-**11** (**Figure 2.3**). Of compounds **1**, **3**, **5**, **6**, and **8**-**11**, only **1** and **6** are found in all the strains profiled (**Figure B.1**). The ΔORF15 mutation increases the amount of compound **6** by approximately tenfold relative to the wild-type, whereas the Δ*hlmI*/ΔORF15 double mutations increase the amount of **6** by two and three orders of magnitude, respectively. The primary enamine-containing compounds **5**, **9**, and **11** were only observed in extracts of the Δ*hlmI*/ΔORF15 mutant.

Activity of Dimethyldihydroholomycin against S. clavuligerus Strains: To evaluate the effect of *S*-methylation on the activity of dihydroholomycin, a *S*,*S*'-dimethyldihydroholomycin standard was chemically synthesized. Attempts to synthesize the *S*-monomethyldihydroholomycin was unsuccessful due to apparent rapid formation of the dimethyl species. Therefore the inhibitory effect of monomethylated compounds were not tested. The ^{1}H and ^{13}C NMR spectroscopic data of the synthetic *S*,*S*'-dimethyldihydroholomycin were consistent with those of the isolated sample of compound **10**, confirming structural assignments (**Figures B.13** and **B.14**).[17] The antimicrobial activity of *S*,*S*'-dimethyldihydroholomycin was tested in a bioassay against *S. clavuligerus* wild-type, ΔORF15, Δ*hlmI*, and Δ*hlmI*/ΔORF15 strains in parallel with the same concentrations of holomycin. *S*,*S*'-dimethyldihydroholomycin did not inhibit the growth of any of the tested strains, whereas holomycin displayed inhibitory effect against all four strains, especially for the Δ*hlmI* mutants (**Figure B.3**) as observed in previous work.[9] These results suggest that the dithiol groups of reduced holomycin are relevant to the biological activity of this antibiotic and *S*-methylation of the dithiols in the producing bacteria significantly decrease that antimicrobial activity as a self-protection mechanism. However, the activity of dihydroholomycin could not be tested directly as it is oxidized in minutes to holomycin in air.[9]

Conclusions: These studies were undertaken with the specific question of how *S. clavuligerus* protects itself during holomycin production when it has insufficient capacity (e.g., in the *hlmI*-null mutant) to oxidize the presumed proximal toxin, dihydroholomycin, to the disulfide that is normally excreted into the medium. More generally the study probes the question of what chemical strategies producer microbes utilize to blunt the reactivity of unwanted thiol groups when disulfide formation is suppressed. Our findings indicate that *S*-methylation in combination with dimerization plays a major role. Building upon previous studies that utilized NMR- and LC/MS-based comparative metabolomics

to connect metabolic outputs to specific biosynthetic pathway gene mutations,[14] we have profiled the chemical species upregulated in the Δ*hlmI*/ΔORF15 mutant relative to ΔORF15. The three Δ*hlmI*/ΔORF15-upregulated metabolites fully characterized in this study-**3**, **5** and **6**-all represent novel secondary metabolites, one of which, the dimeric **6**, is also consistently produced by the wild-type strain. The structures of several additional Δ*hlmI*/ΔORF15-upregulated compounds were proposed with high confidence, although their characterization by NMR spectroscopy remained incomplete due to the characteristically poor line-shapes of these compounds. It was demonstrated that a mix of mono- and dimethyl thioethers, with and without *N*-acetylation, accumulate in the mutant extracts (**Figure 2.4**). From prior in vitro studies HlmA and HlmI were assigned as enzymes catalyzing the last two steps in the holomycin biosynthetic pathway. assembly.[6,9] Results of the current study are consistent with these roles, although either step could be the ultimate step. The unusual dimeric adduct **6** detected at higher levels in the Δ*hlmI*/ΔORF15 mutant, relative to all the other *S. clavuligerus* strains probed, has a thioether bridge from the exocyclic ene-sulfur of one pyrrolone unit to the C-6 position of the *N*-acylaminopyrrolone moiety of a second unit (**Figure 2.2**). In this adduct, two of the sulfurs are *S*-methylated, as found in the monomeric adducts noted above (**3**). One might have expected an intermolecular disulfide from the two remaining thiols (the disulfide can be seen but only after oxidation during workup of crude extracts) but this adduct instead contains a thioether linkage, and thus one of the anticipated sulfur atoms appears to have been lost. One possible path to its formation would be for a monomethylated dihydroholomycin, with a free exocyclic enethiol to add into the C5=C6 double bond of a second molecule through conjugate addition (**Figure 2.4**). The resultant tetrahedral adduct could then eliminate -SH or -SCH_3, depending on the state of the sulfur that is departing, to yield the dimeric thioether.[18] This route illustrates the nucleophilic character of the exocyclic enethiol of reduced holomycin, and demonstrates why the producer organism would go to substantial lengths to avoid its accumulation. It

also reveals the potential electrophilicity of the cyclic enamide. The reactivity of this enamide is presumably enhanced by acylation of the exocyclic amine by HlmA; however, it may also be further augmented by oxidation of one or other of the sulfur atoms, such as in compound **11**.[18]

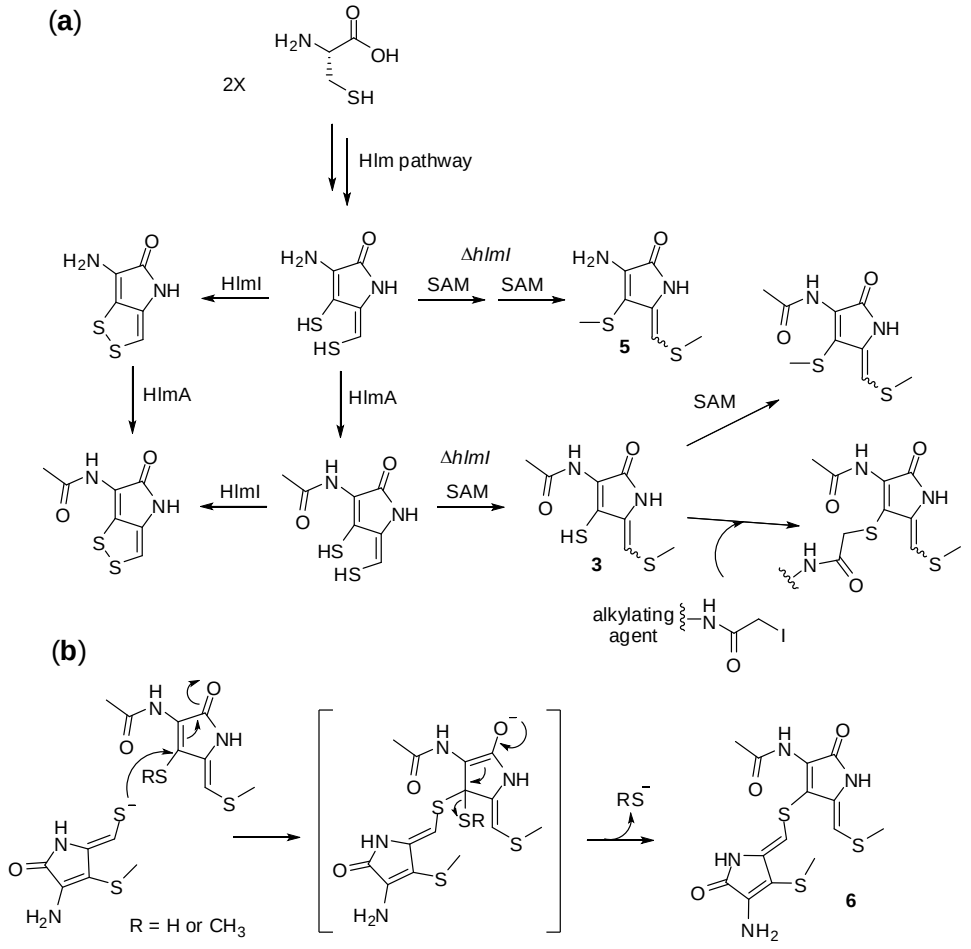


Figure 2.4: Proposed (bio)synthetic pathways. In the Δ*hlmI*/ΔORF15 mutant strain generation of (**a**) mono- and bismethylated intermediates (including **3**, **5**, and **10**) that have been characterized by MS and/or NMR, and (**b**) heterodimeric compound **6**. SAM: *S*-adenosylmethionine.

In particular, these results summarize the biochemical consequence of upregulating holomycin production in *S. clavuligerus* (due to the deletion of the ORF15 gene) when the last-step oxidation is disabled. In absence of the dithiol oxidase HlmI *S.*

clavuligerus is subjected to build-up of the active (reduced) form of its own toxin. One way to blunt this activity is to remove the nucleophilicity of the enethiol groups. While *S*-acetylation possibly provides an alternate detoxification strategy, the resultant thioesters would be readily hydrolysable and rather create a reactive acylating agent. Instead *S*-methylation, likely the result of methyl transfer from *S*-adenosylmethionine (SAM), is observed. This process could be sequential with monomethyl species as free intermediates (**Figure 2.4**). The inactivity of the dimethyldihydroholomycin standard observed in the bioassay supports the role of *S*-methylation as a back-up plan for self-protection. Another possible route of detoxification is for dihydroholomycin to react with mycothiol or mycothione, an abundant cellular thiol in actinomycete,[19] through disulfide exchange or nucleophilic addition. However, mycothiol-holomycin adducts were not detected. The identity of the putative SAM-dependent methyltransferase in *S. clavuligerus* producers is as yet unknown. One might expect an analogous *S*-methylation self-protection strategy in the Δ*gliT* mutants of *A. fumigatus* to blunt the nucleophilicity and consequent toxicity of the dithiol form of gliotoxin. In fact, a wild-type gliotoxin producer, *Gliocladium deliquescens* adopts this thiol methylation strategy, and has been reported to irreversibly methylate gliotoxin to yield bisdethiobis(methylthio)-gliotoxin.[20] Similarly, bismethylthiol forms of gliotoxin have been isolated from other fungal strains.[21,22] These are exactly the metabolites that should arise from bismethylation of the dihydro form of gliotoxin in detoxification routes. The bisthioether versions of gliotoxin have been shown to be substantially less toxic than gliotoxin itself.[21,23] Additionally, as described in the DANS-based study described in Chapter 1, both mono- and bismethylation species were detected on incompletely processed scaffolds.[14] The *S*-methylation may be catalyzed by the methyltransferase *GliN* encoded as a backup plan to protect the producer from the toxicity of the biosynthetic intermediates. *S*-methylation of dithiols in bacterial and fungal strains, using the readily available SAM as an electrophilic methyl donor, may thus be the preferred protection

mechanism against endogenous and exogenous dithiol forms of toxic metabolites when the native dithiol oxidase is absent. Indirectly, this study strengthens the hypothesis that the molecular species with antibiotic properties is the penultimate metabolite in the holomycin pathway, dihydroholomycin. In the producing organism, this metabolite is rapidly converted to the bicyclic disulfide through HlmI action and then exported to the extracellular matrix. When taken up by a susceptible neighboring microbe, the disulfide will be in a microenvironment rich in thiols. Reduction to the mixed disulfide and/or the dithiol in a target cell may be nonenzymatic or enzymatic and generate the reactive nucleophilic enethiol forms of the antibiotic. It is possible that dihydroholomycin, like the dihydro form of gliotoxin, exerts some of its toxicity through redox cycling to generate reactive oxygen species, in which case *S*-methylation can also prevent that toxicity. Although inhibition of RNA synthesis has been proposed as an end effect of the holomycin class of antibiotics[2,3] the molecular targets of this compact yet latentlyreactive framework remain to be fully characterized.

REFERENCES

(1) Jimenez, A.; Tipper, D. J.; Davies, J. *Antimicrob Agents Ch* **1973**, *3*, 729.

(2) Tipper, D. J. *J Bacteriol* **1973**, *116*, 245.

(3) Oliva, B.; O'Neill, A.; Wilson, J. M.; O'Hanlon, P. J.; Chopra, I. *Antimicrob Agents Ch* **2001**, *45*, 532.

(4) Gardiner, D. M.; Waring, P.; Howlett, B. J. *Microbiology* **2005**, *151*, 1021.

(5) Furumai, R.; Matsuyama, A.; Kobashi, N.; Lee, K. H.; Nishiyama, N.; Nakajima, I.; Tanaka, A.; Komatsu, Y.; Nishino, N.; Yoshida, M.; Horinouchi, S. *Cancer Research* **2002**, *62*, 4916.

(6) Li, B.; Walsh, C. T. *PNAS USA* **2010**, *107*, 19731.

(7) Schrettl, M.; Carberry, S.; Kavanagh, K.; Haas, H.; Jones, G. W.; O'Brien, J.; Nolan, A.; Stephens, J.; Fenelon, O.; Doyle, S. *PLoS Pathog* **2010**, *6*, e1000952.

(8) Waring, P.; Sjaarda, A.; Lin, Q. H. *Biochem Pharm* **1995**, *49*, 1195.

(9) Li, B.; Walsh, C. T. *Biochemistry* **2011**, *50*, 4615.

(10) de la Fuente, A.; Lorenzana, L. M.; Martin, J. F.; Liras, P. *J Bacteriol* **2002**, *184*, 6559.

(11) Celmer, W. D.; Solomons, I. A. *J Am Chem Soc* **1955**, *77*, 2861.

(12) Okamura, K.; Soga, K.; Shimauchi, Y.; Ishikura, T. *J Antibiot* **1977**, *30*, 334.

(13) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *PNAS USA* **2007**, *104*, 1506.

(14) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(15) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *PNAS USA* **2009**, *106*, 7708.

(16) Davies, J. *Curr Opin Chem Bio* **2011**, *15*, 5.

(17) Hjelmgaard, T.; Givskov, M.; Nielsen, J. *Org Biomole Chem* **2007**, *5*, 344.

(18) Schachtner, J. E.; Nienaber, J.; Stachel, H. D.; Waisser, K. *Pharmazie* **1999**, *54*, 335.

(19) Newton, G. L.; Buchmeier, N.; Fahey, R. C. *Microbiol Mol Biol R* **2008**, *72*, 471.

(20) Kirby, G. W.; Robins, D. J.; Sefton, M. A.; Talekar, R. R. *J Chem Soc Perk T 1* **1980**, 119.

(21) Li, X. F.; Kim, S. K.; Nam, K. W.; Kang, J. S.; Choi, H. D.; Son, B. W. *J Antibiot* **2006**, *59*, 248.

(22) Lee, H. J.; Lee, J. H.; Hwang, B. Y.; Kim, H. S.; Lee, J. J. *Arch Pharm Res* **2001**, *24*, 397.

(23) Sun, Y.; Takada, K.; Takemoto, Y.; Yoshida, M.; Nogi, Y.; Okada, S.; Matsunaga, S. *J Nat Prod* **2012**, *75*, 111.

CHAPTER 3

HOMOLOGOUS NON-CANONICAL NRPS GENE CLUSTERS MEDIATE REDUNDANT SMALL-MOLECULE BIOSYNTHESIS IN *ASPERGILLUS FLAVUS*

Abstract: In the opportunistic human pathogen *Aspergillus flavus*, two highly homologous orphan gene clusters, *lna* and *lnb*, are regulated by LaeA, a global regulator of secondary metabolism, morphogenesis, and virulence. The *lna* and *lnb* clusters contain two non-canonical non-ribosomal peptide synthetase (NRPS) genes, *lnaA* and *lnbA*, that both lack a conventional condensation domain and share homology to reductases participating in fungal L-lysine biosynthesis. By employing 2D NMR-based comparative metabolomic analyses of knock-out, overexpression, and knock-down strains of *lnaA* and *lnbA*, this study demonstrates that the two clusters encode enzymes that produce partially overlapping sets of metabolites, including previously undescribed piperazines, pyrazines, and morpholines. These results provide evidence for functional redundancy of the *lna* and *lnb* clusters and suggest that one primary function of the NRPSs LnaA and LnbA is the reduction of their substrate, L-tyrosine. Additionally, morphological studies demonstrate that mutations of *lna* pathway genes interfere with sclerotia formation patterns and that one of the identified *lnaA*- and *lnbA*-dependent metabolites strongly affects expression of *lnaB*, a putative reductase/epimerase, suggesting that the *lna* and *lnb* pathways are part of a signaling network regulating *A. flavus* sclerotia formation.

Introduction: Fungi are among the most prolific sources of pharmacologically relevant natural products.[1] This large diversity of fungal small molecules serves important functions in fungal ecology, for example as virulence factors or as chemical defense

agents. However, only a fraction of the biosynthetic capabilities suggested by genomic analyses has been observed under laboratory conditions, because expression of many, perhaps even most biosynthetic pathways depends strongly on environmental conditions.[2-4] This study demonstrate the use of comparative metabolomics[5-8] for the analysis of specific secondary metabolic gene knockout, overexpression and knockdown strains to determine the metabolic output of two non-ribosomal peptide synthetase (NRPS) gene clusters in the ascomycete fungus *Aspergillus flavus,* a crop contaminant[9,10] and opportunistic pathogen, causing aspergillosis in immunocompromised humans.[11,12]

Although the *A. flavus* genome encodes at least 25 polyketide synthase (PKS), 18 NRPS, and two hybrid NRPS-PKS gene clusters,[13] assignments have been made for only four metabolites.[14] Many *A. flavus* secondary metabolite pathways are under the control of the protein LaeA, a global regulator of morphogenesis and virulence factor in *A. flavus* and other pathogenic fungi.[15] Two LaeA-regulated clusters, which were named *lna* and *lnb*, exhibit a striking level of genetic similarity (**Figure 3.1**, **Table C.1**). The *lna* and *lnb* clusters contain two non-canonical NRPS genes with high sequence homology (58% identical at the amino acid level), *lnaA* and *lnbA*, respectively, which are accompanied by matching sets of genes likely coding for tailoring enzymes. *lna* and *lnb* are orphan clusters that have no known associated metabolites and belong to a family of NRPS genes that consist of an adenylation (A) domain, a peptidyl carrier protein (PCP) domain, and a thioester reductase (R) domain, but lack a canonical condensation (C) domain (**Figure 3.1**). The functions of this unusual family of NRPSs, which share homology to reductases participating in fungal L-lysine biosynthesis,[16] have not been explored. In the studies described herein, it is demonstrated that the *lna* and *lnb* clusters encode sets of enzymes that produce overlapping sets of previously undescribed metabolites, and one primary function of the non-canonical NRPSs LnaA and LnbA likely consists in reduction of L-tyrosine. Furthermore, the *lna* and *lnb* biosynthetic

pathways appear to be part of a signaling network that controls the formation of sclerotia, a resilient overwintering structure.

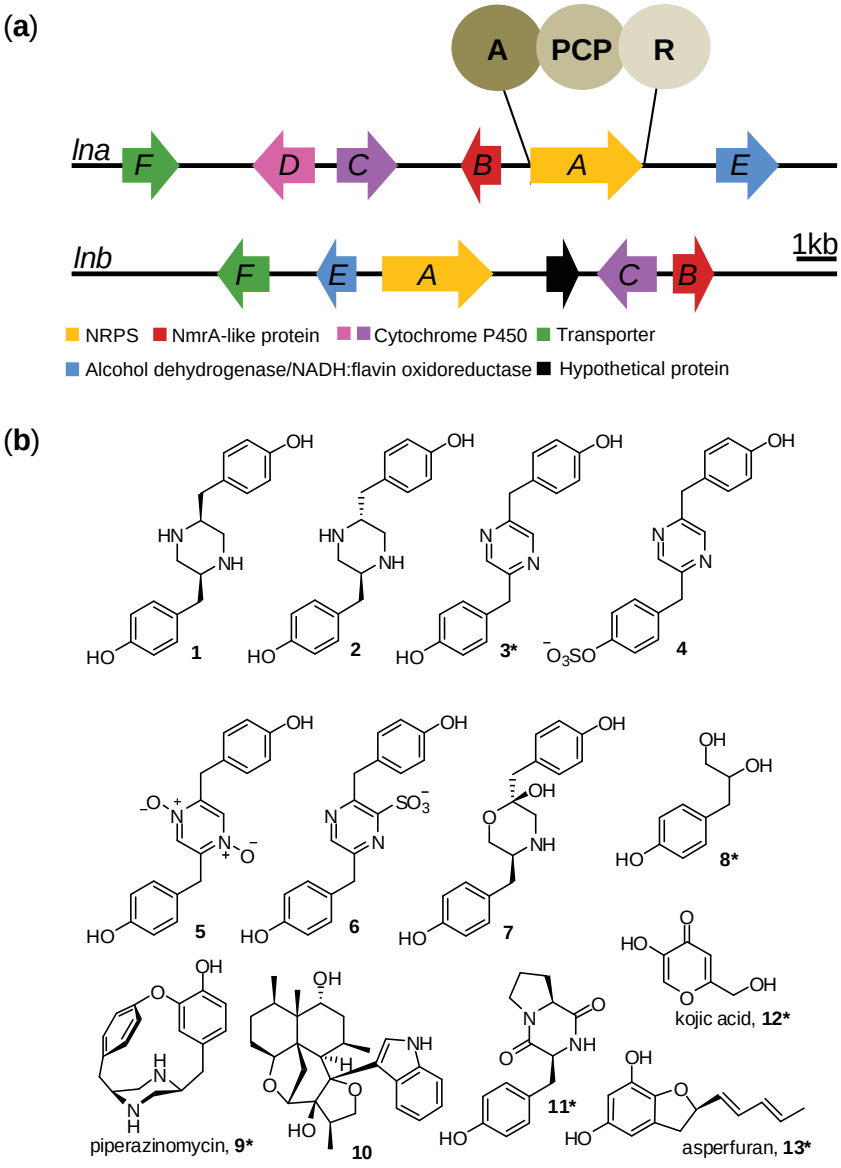


Figure 3.1: *lna* and *lnb* gene clusters and characterized *A. flavus* compounds. (**a**) *A. flavus lna* and *lnb* gene clusters. Both *lna* and *lnb* include non-canonical NRPS genes including adenylation (A), carrier protein (PCP), and a short-chain dehydrogenase/reductase domain (R-domain). There is no *lnaD* analog in the *lnb* gene cluster. See also Appendix C, **Table C.1**. (**b**) Metabolites identified in this study and the known piperazinomycin (**9**). Compounds marked with an asterisk* have been previously described.

Comparative Metabolomics Reveals lna-Dependent Metabolites: To identify *lna*-associated metabolites *via* comparative metabolomics, deletion and overexpression[17] mutants were created for NRPS gene *lnaA* (*ΔlnaA* and *OE::lnaA* [two isolates obtained], respectively) as well as two double mutant strains *ΔlnaA, KD::lnbA* and *OE::lnaA, KD::lnaB*, where KD indicates a knockdown of gene expression using RNAi (strains are listed in **Table C.3**). The overexpression of the NRPS gene *lnaA* was chosen as the primary method of *lna* pathway upregulation because the *lna* and *lnb* clusters lack apparent *cis*-acting regulatory genes, advantageous genetic targets for promoting cluster expression.[18] This strategy has been adopted previously to increase titers of penicillin by overexpression of its NRPS gene *acvA* in *A. nidulans*.[19,20] Unexpectedly, in the *ΔlnaA, KD::lnbA* double mutant strain, in which both the *lna* and *lnb* pathways are disrupted, formation of sclerotia was strongly suppressed relative to wild-type (see Appendix C, **Figure C.1**).

To compare the metabolomes of the wild-type, Δ*lnaA* and O*E::lnaA* strains, differential analysis by 2D NMR spectroscopy (DANS) was employed.[6,8,21] dqfCOSY spectra acquired for crude and pre-fractionated wild-type, Δ*lnaA*, and *OE::lnaA* (both isolates) extracts were compared pair-wise using an overlay algorithm designed to highlight strongly up- or downregulated signals.[8] Whereas comparison of the 2D NMR spectra of the two different *OE::lnaA* strains showed no differential signals, comparison of the *OE::lnaA* spectra with either wild-type or Δ*lnaA* revealed a large number of spin systems present only in the *OE::lnaA* spectra (**Figure 3.2**). The most prominent *OE::lnaA*-upregulated spin system showed a singlet at ~4.5 ppm with long range-couplings (< 1 Hz) to an aromatic AA'BB'-spin system at 6.7 and 7.1 ppm as well as long-range coupling to a singlet at 8.4 ppm. These data suggested a chemical structure featuring a p-hydroxy substituted-benzyl group bound to an *N*-heteroaromatic system. In addition, *OE::lnaA*-specific dqfCOSY signals were observed indicating the presence of at least three structurally related compounds as well as a second group of at least four

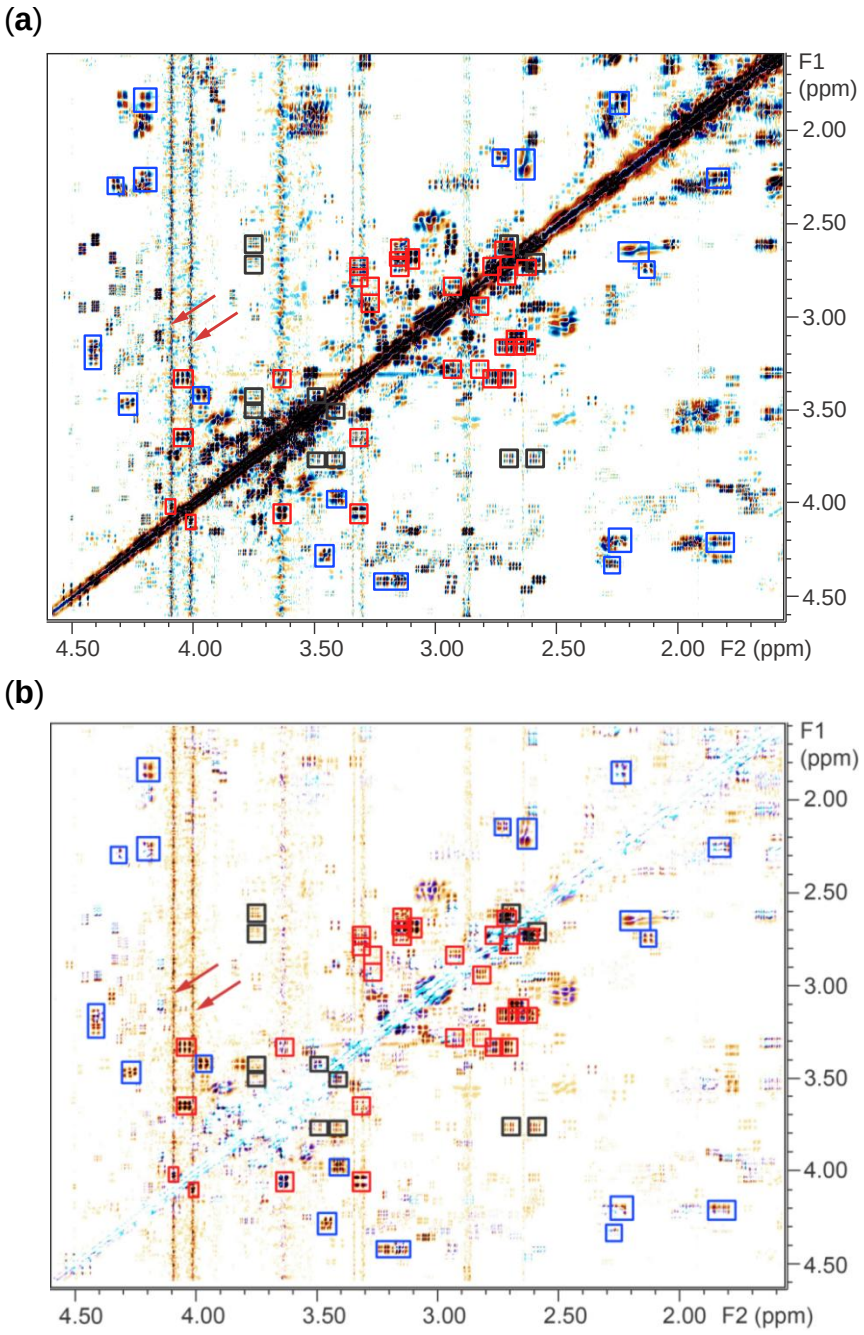


Figure 3.2: Pool 3 *OE::lnaA* and DANS spectrum. (**a**) dqfCOSY spectrum of *OE::lnaA* metabolite fraction used for characterization of *lnaA*-dependent compounds detected *via* DANS. (**b**) Section of the pool-3 DANS overlay of the *OE::lnaA* dqfCOSY spectrum with the Δ*lnaA* dqfCOSY spectrum. For both panels, cross peaks marked with red boxes or arrows represent novel *lnaA*-dependent compounds. Black boxes mark the known compound **8** that is upregulated in *OE::lnaA*. Cross peaks marked blue represent species that were not reproducibly differential or were suppressed incompletely in the overlay due to chemical shift variation.

different compounds, all of which appeared to include a 3-(p-hydroxyphenyl)-substituted 1,2-diaminopropane- or 2-amino-1-hydroxypropane-derived moiety. Using available structure databases, it was determined that the *OE::lnaA*-upregulated partial structures did not match any known metabolites from Aspergilli or other fungi.

For further characterization of the *lnaA*-dependent metabolites, scaled-up *OE::lnaA* cultures (TSA7.51) were fractionated using reverse-phase chromatography. Comparison of ^{1}H and dqfCOSY spectra of crude and fractionated extracts indicated that the *OE::lnaA*-upregulated compounds were not altered by chromatography. Using DANS-detected diagnostic NMR signals, fractions containing the *OE::lnaA*-upregulated compounds were selected for further analysis. For one major *OE::lnaA*-upregulated compound, ^{1}H,^{13}C-HMBC and ^{1}H,^{13}C-HSQC spectra (for NMR spectroscopic data, see **Tables C.4**-**C.10**) were consistent with an *O*-substituted derivative of the known actinopolymorphol (**3**, **Figure 3.1**).[22] Chemical shift values of protons and carbons of one of the aromatic rings suggested *O*-sulfonation[23] and thus a molecular formula of $C_{18}H_{16}N_2O_5S$ for this compound (**4**, **Figure 3.1**), which was confirmed by positive electrospray ionization- ($ESI^{+/-}$) high-resolution- (HR-) MS (Appendix C, section 13). Compound **4** was accompanied by smaller amounts of compound **3** and two additional highly polar derivatives of **3**. The NMR spectroscopic data of one of these compounds suggested *N*-oxidation, which was confirmed by UPLC/ESI^{+}-HRMS analysis indicating a molecular formula of $C_{18}H_{16}N_2O_4$, thus revealing an unusual *N*,*N*-dioxide, (for a structurally related synthetic compound, see[24]). NMR spectroscopic data of the second derivative of compound **3** indicated substitution at C-3. UPLC/$ESI^{+/-}$-HRMS spectra revealed a molecular formula of $C_{18}H_{16}N_2O_5S$ (Appendix C, section 13), indicating C-sulfonylation in this metabolite, compound **6**.

Analysis of fractions containing the putative (p-hydroxyphenyl)-substituted 1,2-diaminopropane- or 2-amino-1-hydroxypropane derivatives revealed two compounds with very similar ^{1}H chemical shift and coupling constant values. dqfCOSY and ^{1}H,^{13}C-

HMBC spectra confirmed that these compounds incorporate 3-(p-hydroxyphenyl)-substituted 1,2-diaminopropane moieties. HPLC/ESI$^{+/-}$-MS and UPLC/ESI$^{+/-}$-HRMS analyses indicated that both analytes have a molecular formula of $C_{18}H_{22}N_2O_2$. Taken together; these spectroscopic data suggested the pair of diastereomeric piperazines **1** and **2**. Comparison with synthetic samples prepared *via* cyclo(L-Tyr-L-Tyr) and cyclo(D-Tyr-L-Tyr) derivatives confirmed these assignments (Appendix C, section 8). The structures of **1** and **2** are closely related to piperazinomycin (**9**),[25] a bacterial metabolite first identified from *Streptomyces olivoreticuli*, which may be derived from oxidative macrocyclization of **1**-like precursors.

Another (p-hydroxyphenyl)-substituted propane derivative included a quaternary carbon at 93.1 ppm as well as two isolated methylene groups and a CH_2O-CH-CH_2-phenyl moiety. ESI$^+$-HRMS demonstrated a molecular formula of $C_{18}H_{21}NO_4$, which in conjunction with the dqfCOSY and ^{1}H,^{13}C-HMBC data led to identification of the hemiacetal-containing morpholine **7**. ROESY correlations between the hydroxy-proton with the axial proton at C-6 defined the relative configuration of **7** as (*2R**,*5S**). HPLC/ESI$^+$-MS analysis of samples containing **7** showed two similarly large chromatographic peaks with mass spectra and fragmentation pattern consistent with the structure of **7**, suggesting that under the HPLC conditions used **7** undergoes epimerization. The piperazines **1** and **2** and morpholine **7** are accompanied by 3-(p-hydroxyphenyl)-1,2-propanediol (**8**), a previously described fungal metabolite[26].

All described *OE::lnaA*-upregulated compounds (**1**-**8**) were consistently present in *OE::lnaA* rich media cultures and undetectable by NMR spectroscopy in WT and Δ*lnaA* rich media cultures. Of the *OE::lnaA*-upregulated compounds characterized, seven represent novel metabolites (**1**, **2**, and **4**-**7**) and two have not previously been reported from *A. flavus* (**3** and **8**). In addition, comparative metabolomic analyses occasional showed the production of several other metabolites (**11**-**13**) whose structures appeared unrelated to those of **1**-**8**, suggesting that they are not directly derived from

lna cluster genes. In one set of cultures (liquid stationary GMM cultures), DANS analysis revealed five spin systems in wild-type and *OE::lnaA* extracts that were absent in Δ*lnaA.* Further analysis suggested one 4-proton spin system representing an indole moiety, whereas the other four spin systems appeared to belong to an oxygenated aliphatic ring system including several methyl doublets. A search of the established partial structures using Reaxys and SciFinder suggested that they represent undescribed chemical structure(s) perhaps related to a family of previously described idole-containing *A. flavus* compound, the aflavinines[27].

The analyte of interest was enriched using semi-preparative HPLC applied to combined wild-type and *OE::lnaA* extracts (Appendix C, section 9). 2D NMR analysis of the enriched sample using $^{1}H,^{13}C$-gHMBCAD, $^{1}H,^{13}C$-HSQC, and dqfCOSY supported the structure of **10** (**Figure 3.1**). Similar to the aflavinines, compound **10** showed strong 3-bond $^{1}H,^{13}C$-correlations from the methyl protons of C-26, 27, 28 and 29 to several key ^{13}C atoms establishing the oligocyclic carbon skeleton. Additionally, $^{1}H,^{13}C$-gHMBCAD correlations from the protons of C-12, 16, 23, and 28 to the central tertiary C-14 helped connect the several contiguous $^{1}H,^{1}H$-spin systems observed in the dqfCOSY to each other (**Figure 3.3**). ROESY spectra further confirmed the molecular architecture of **10**, demonstrating that the C-2 proton of the indole moiety is positioned proximal to the C-26 and C-29 methyl groups. The positioning of the indole further confirms the relative stereo chemistry of C-10 and C-11. ROESY correlations between the C-15 proton with the C-18 proton suggested that the 6-membered ring formed by C-15 through C-19 is in a boat conformation. The adjacent 6-membered ring formed by C-19 through C-23 and C-14 shows *J*-coupling values and ROESY correlations consistent with a chair conformation (**Figure 3.3**). These 2D NMR spectroscopic data in combination with HRMS data support the structural assignment of **10** (**Table C.11**).

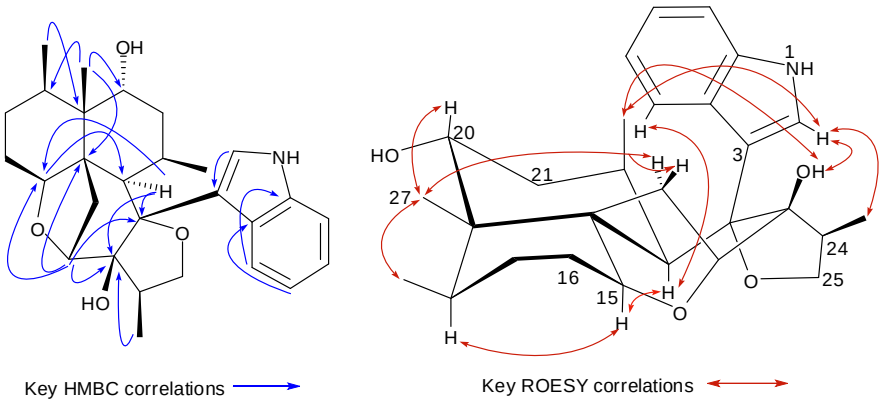


Figure 3.3: Diagnostic structural correlations for compound 10. Key HMBC correlations are displayed as blue signally headed arrows. Key through space ROESY correlations are displayed as red double headed arrows.

LC/ESI-MS Profiling of the OE::lnaA-upregulated Metabolites in the Wild-type, ΔlnaA, and ΔlnaA, KD::lnbA A. flavus Strains: Given that compounds **1**-**8** were produced when *lnaA* was overexpressed, studies were undertaken to determine if **1**-**8** production can be elicited in wild-type *A. flavus* under specific culture conditions, and whether these metabolites are in fact strictly *lna*-dependent (see **Table 3.1**). Highly sensitive single ion-monitoring MS (SIMMS) analysis of *ΔlnaA*, wild-type, and *OE::lnaA* extracts derived from cultures grown using conditions previously shown to support *lnaA* expression,[4] revealed the presence of piperazines **1** and **2** in WT, although at ~100-fold lower concentrations than *OE::lnaA* (see Appendix C, section 11). Pyrazine **3** was found only in some, but not all wild-type extracts, whereas **4**-**8** were not detected. *OE::lnaA* produced compounds **1**-**3**, **7** and **8**, whereas **4**-**6** were absent under these culturing conditions. Notably, SIMMS also detected very small quantities of **1** and **2** in *ΔlnaA* extracts, at levels roughly 10 times lower than in wild-type. This suggested that **1** and **2** are perhaps also produced by enzymes encoded by the homologous *lnb* cluster. To test this hypothesis, extracts derived from two different *ΔlnaA, KD::lnbA* double mutant strains were analyzed by HPLC/ESI$^+$-SIMMS, which showed that compounds **1**-**8** are

completely absent in the double mutants, whereas wild-type and *ΔlnaA*, grown as positive controls in parallel, produced **1** and **2** as before (**Figure C.3**). These results indicate that the amounts of **1** and **2** found in *ΔlnaA* are derived from *lnbA* expression and thus the *lna* and *lnb* clusters (**Figure 3.1**) encode partially redundant biosynthetic pathways.

Compound	*OE::lnaA* (rich media)	*OE::lnaA*	*OE::lnaA, KD::lnaB*	Wild type	Δ*lnaA*	Δ*lnaA, KDlnbA*
1	+++	+++	+++	++	+	**-**
2	+++	+++	+++	++	+	**-**
3	+++	+++	+++	++*	**-**	**-**
4	+++	**-**	-	**-**	**-**	**-**
5	+++	**-**	-	**-**	**-**	**-**
6	+++	**-**	-	**-**	**-**	**-**
7	+++	+++	+++	**-**	**-**	**-**
8	+++	+++	+++	**-**	**-**	**-**

Table 3.1: *A. flavus* strains and occurrence of *lna/lnb*-associated metabolites 1-8. "+++" = abundant, detected by DANS and MS, "++" = detected by MS only ("*" = occasionally detected), and "+" = detected by MS, ~10 fold less abundant than in WT. Strains were cultured on GMM unless indicated otherwise.

Determination of LnaA and LnbA Substrate Specificity: Given that the structures of **1**-**8** appear to be tyrosine (Tyr) derived, the two putative NRPS proteins LnaA and LnbA were tested for their ability to specifically activate D/L-Tyr by assaying amino acid-dependent ATP-[^{32}P]-pyrophosphate exchange activity.[28] Both LnaA and LnbA specifically activated L-Tyr, although LnaA was somewhat D/L-unspecific and also activated D-Tyr (**Figure 3.4**). However, neither *lnaA* nor *lnbA* include a condensation domain, and correspondingly, none of the *lna*- or *lnb*-associated metabolites identified in the wild-type nor *OE::lnaA* background feature a peptide bond. In addition, neither protein activated the L-Tyr-L-Tyr dipeptide (**Figure C.4**). However, both *lnaA* and *lnbA* include C-terminal R- (putative short-chain reductase and/or epimerase) domains, with

amino acid sequence similarity to other microbial reductase domains.[29,30] Based on the chemical structures of the identified *lna*- and *lnb*-dependent metabolites, it appeared likely that the R-domains in LnaA and LnbA are involved in reduction of L-Tyr or L-Tyr derivatives. In a biosynthetic model, LnaA and/or LnaB-derived L-Tyr aldehyde (**15**) may form a dimer such as **17**, perhaps *via* a tethered intermediate, e.g. **16** (**Figure 3.5**).

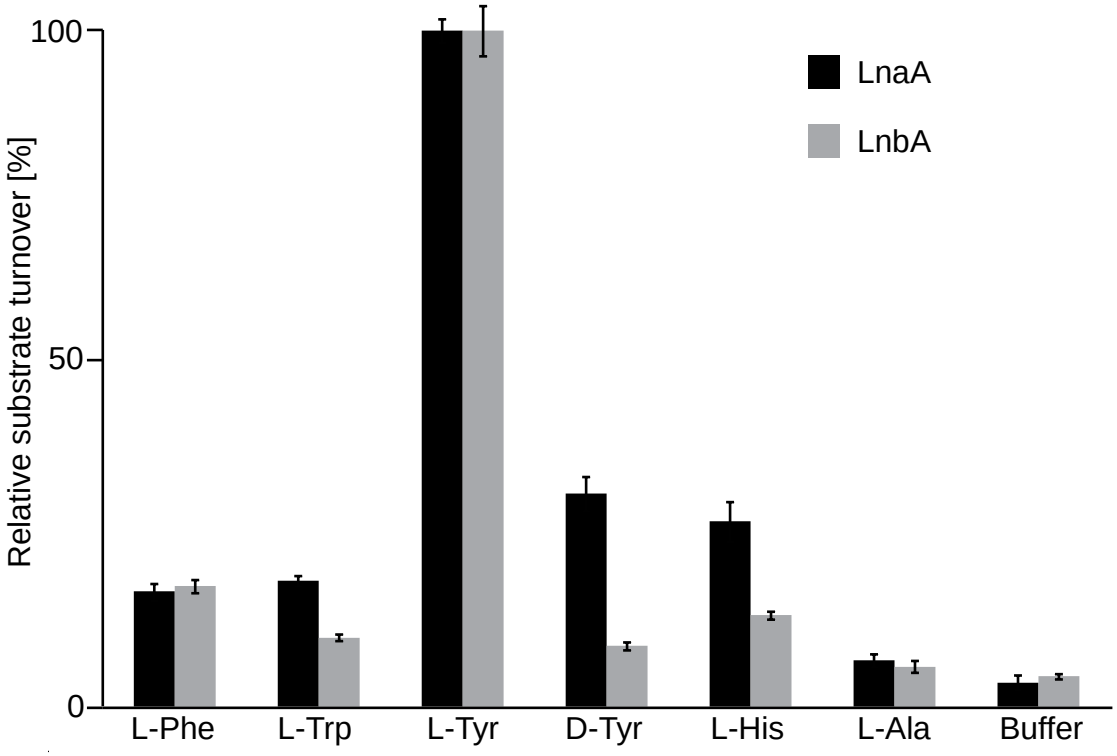


Figure 3.4: Substrate specificity of recombinant LnaA and LnbA. LnaA (black bars) and LnbA (grey). were tested for their ability to adenylate various amino acids. Shown are relative activity values normalized to L-Tyr activity. Error bars, SD.

2D NMR and LC/ESI-MS-based Metabolomic Study of the Putative Reductase Encoding Gene lnaB: Formation of the most abundant *lna*- or *lnb*-associated metabolites requires additional reduction from aldehydes or imines to corresponding

amines and alcohols. These reductive steps seemed likely to involve putative NmrA-like proteins encoded by *lnaB* or *lnbB* (**Figure 3.1**), as NmrA-like proteins are believed to serve enzymatic functions as epimerases or reductases,[31] in addition to possible roles in gene regulation.[19]

To examine the putative role of LnaB in the biosynthesis of **1** and **2**, the metabolite profile of an *OE::lnaA, KD::lnaB* double mutant strain was compared to that of *OE::lnaA* by 2D NMR and HPLC/ESI$^+$-SIMMS. These analyses showed that the morpholine **7** is much more abundant in *OE::lnaA, KD::lnaB* than in the *OE::lnaA* background, relative to the piperazines **1** and **2** (**Figure 3.5**). This shift towards a relatively greater production of morpholine **7**, whose biosynthesis requires one reduction step less than production of **1** and **2**, suggests that LnaB participates in the reduction of LnaA-derived intermediates. Knockdown of LnaB would result in increased accumulation of **17**, which may partially undergo hydrolysis leading to increased production of **7** (**Figures 3.5** & **C.5**). Notably, production of piperazines **1** and **2** is not fully abolished in the *OE::lnaA, KD::lnaB* strain, and formation of morpholine **7** from L-Tyr aldehyde (**15**) or **17** still requires one reductive step. The residual reductase activity in the *OE::lnaA, KD::lnaB* double mutant could result from expression of the second *nmrA*-like gene *lnbB* (or another related gene proximal to *lnbB*, **Table C.1**) or incomplete silencing of *lnaB*. Lastly, whereas cyclization of **17** to a corresponding diimine followed by LnaB- (or LnbB-) mediated reduction leads to piperazines **1** and **2**, oxidation, non-enzymatically or catalyzed by putative oxidases LnaC, LnaD, or LnbC, would explain formation of pyrazine **3**.

HPLC/ESI$^+$-SIMMS further showed that the ratio of the two diastereomers **1** and **2**, is significantly greater in wild-type than in *ΔlnaA* (**Figure 3.6** and **Figure C.6**), suggesting that the *lnb* pathway produces relatively larger amounts of the (*2R**,*5S**)-isomer, **2**, than the *lna* pathway. Given that LnbA activates L-Tyr with high selectivity, it seems unlikely that this greater relative abundance of **2** results from incorporation of

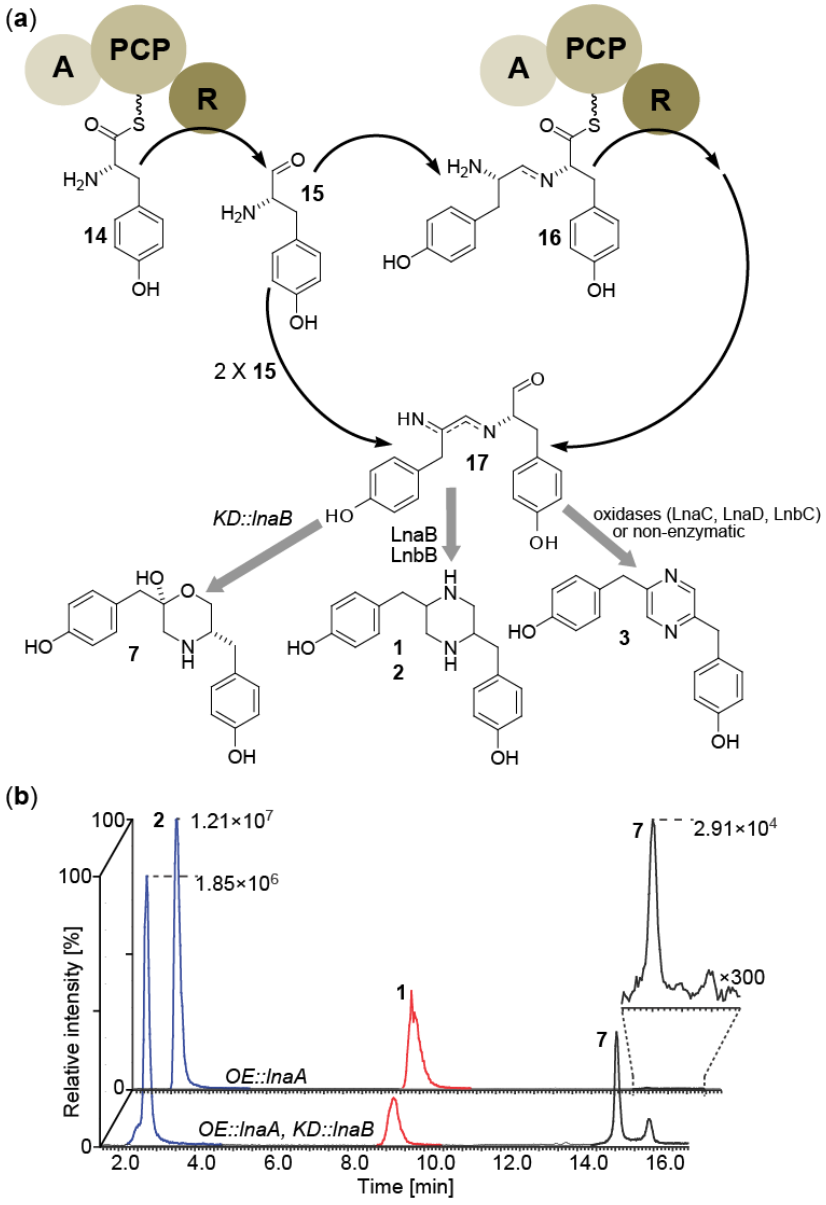


Figure 3.5: Simplified biosynthetic model for compounds 1, 2, 3, and 7 and ion chromatogram comparison of *A. flavus* strains *OE::lnaA* and *OE::lnaA, KD::lnaB*. (**a**) Proposed biosynthesis of *lnaA*-dependent metabolites (see also **Figure C.6**). Tethered L-Tyr (**14**) is reduced to **15** by R-domains of LnaA or LnbA, resulting in formation of a dimeric imine (**17**) potentially *via* a tethered intermediate (**16**). Cyclization and reduction by LnaB (and perhaps LnbB) leads to formation of **1** or **2**, whereas oxidation yields **3**. In the absence of LnaB, isomerization of **17** (or a related intermediate) to the enamine followed by loss of ammonia leads to increased formation of **7**. (**b**) HPLC/ESI$^+$-SIMMS analysis corresponding to **1**, **2**, and **7** for the *OE::lnaA* and *OE::lnaA, KD::lnaB* strains. The *OE::lnaA, KD::lnaB* strain shows a dramatic increase in production of **7** relative to **1** and **2**, compared to the *OE::lnaA* strain.

both D-Tyr and L-Tyr. Instead, these stereochemical differences probably originate from different degrees of epimerization at the aldehyde or imine stage. Next, it was asked asked whether the ratio of diastereomers is actively regulated at the gene expression level. Northern analysis of wild-type cultures incubated with synthetic **1** or **2** showed that treatment with **2** did not lead to any significant changes in *lna* gene expression, whereas addition of **1** to wild-type cultures greatly increased expression of *lnaB* (**Figure 3.6**). This result strongly suggests metabolite-mediated crosstalk between the *lna* and *lnb* pathways.

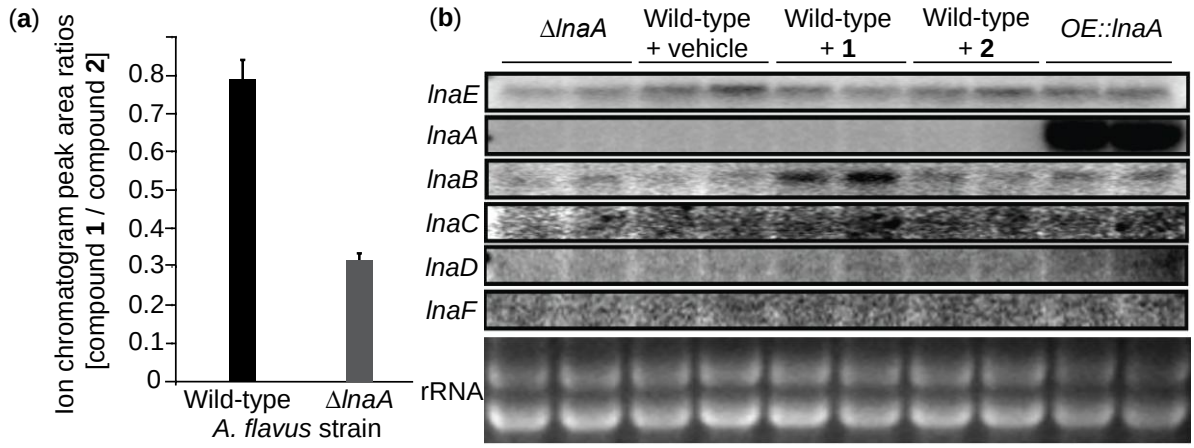


Figure 3.6: The Δ*lnaA* mutation modulates the compound 1 / compound 2 ratio and the influence of compound 1 on lna cluster expression. (**a**) Comparison of the diastereomeric ratios of compounds **1** and **2** in WT and *ΔlnaA* metabolite extracts. Error bars, 1 SD. (**b**) Effect of supplementation with synthetic **1** or **2** on *lna* gene cluster expression. Northern analysis showed that addition of **1**, but not **2** increased *lnaB* expression. rRNA served as loading control.

Conclusions: Employing comparative metabolomics with appropriate *lna* cluster mutations revealed eight *lna*-associated metabolites, **1**-**8**, of which **1**-**3** could be detected in wild-type *A. flavus*. The two most consistently produced *lna* metabolites, **1** and **2**, are also produced under participation of *lnb* genes, providing a first example for partially redundant biosynthesis involving an NRPS-like pathway. LnaA and LnbA

deviate from canonical NRPS domain structures in that they lack condensation domains, and thus, as the structures of identified metabolites suggest, do not act as peptide synthetases. Instead, LnaA and LnbA follow the domain layout of Lys2 and related fungal enzymes that serve as α-aminoadipate semialdehyde reductases in fungal L-lysine biosynthesis (24% and 25% identity of LnaA and LnaB to *A. flavus* Lys2, respectively[16]). Examination of fungal genomes showed that LnaA-like proteins are primarily confined to the Ascomycete order Onygenales, with 1-5 different LnaA-like proteins in different Aspergilli. Recently published genomes (e.g., of *Serpula lacrymans* and *Heterobasidion annosum*[32,33]) show that LnaA/LnbA-type enzymes are also encoded in both basidiomycete genomes. The piperazinomycin-producing bacteria (*Streptomyces* sp.) have not been sequenced; however, available *Streptomyces* genomes contain a number of putative genes encoding proteins with high (up to 38%) amino acid sequence similarity to LnaA. The close similarity of some bacterial NRPS-like proteins to LnaA/LnaB despite bacteria not synthesizing lysine *via* the Lys2 pathway suggests a complex evolution of these proteins[34].

This work provides the first evidence that Lys2-type enzymes also function in secondary metabolism and additionally suggests that NRPS-like proteins may serve as amino-acid reductases, indicating that a reaction hitherto considered as an NRPS-offloading mode can stand by itself. This study also demonstrates the consequences of *lnaA*/*lnbA* gene disruption/silencing and the concomitant loss of piperazine/pyrazine/morpholine secondary products on fungal development, linking sclerotia formation in *A. flavus* and the *lna*/*lnb* metabolic pathways. This observation adds a unique role to previously known functions of NRPS genes in fungi. The occurrence of a functionally duplicated biosynthetic pathway can be interpreted as a safeguard to ensure timely sclerotial production and hence, persistence during unfavorable environmental conditions. The results of this study are suggestive of a complex signaling network regulating biosynthesis of the *lna*- and *lnb*-pathway

metabolites, which may include cross-pathway interactions mediated by sensing of **1** and **2** or shared biosynthetic intermediates. Detailed analysis of the metabolomes and associated phenotypes of *lna*/*lnb* single and double knock-out strains will be required to clarify the extent of interactions between the *lna*/*lnb* clusters and determine the roles of the identified metabolites for sclerotia formation and other aspects of *A. flavus* biology.

***Follow Up Investigation of* Δ*lnaA, KD::lnbA (TSA 22.1)*:** During the *A. flavus lna*/*lnb* gene cluster studies, several interesting observations were made relating to the Δ*lnaA, KD::lnbA* strain (TSA 22.1). It has been demonstrated previously that wild-type *A. flavus* shows less sclerotial production when grown in the light compared to cultures incubated in the dark. In the course of studying TSA 22.1 it was observed that this double mutant strain shows the opposite behavior, producing many sclerotia in the light and few to no sclerotia in the dark (**Figure 3.7**). Several studies have described regulatory pathways that are affected by the absence or presence of light, including the global regulator of secondary metabolisms LaeA[35] and the related VelB-VeA-LaeA complex.[36] This association directly implicates secondary metabolic regulation and pathways that respond to light. Given the observed phenotypes of TSA 22.1 and the established relationship between secondary metabolite production under light/dark conditions, a comparative metabolomic study was undertaken comparing wild-type and TSA 22.1 cultured in the light/dark (wild-type-light/dark and TSA 22.1-light/dark respectively).

Initial comparative HPLC/ESI$^+$-MS analysis of extracts derived from wild-type-light and TSA 22.1-light cultures showed an ion with *m*/z 241.2 dramatically upregulated in TSA 22.1-light (~12 fold) relative to wild-type-light. The *m*/z 241.2 ion was accompanied by an ion *m*/z 543.2 with the same retention time and ion chromatogram peak shape as *m*/z 241.2. A search for know *A. flavus* metabolites with the mass 240.1 amu (assuming *m*/z 241.2 represent a $[M+H]^+$ ion) resulted in one compound, hydroxy aspergillic acid (**18**, **Figure 3.8**). Supporting the hypothesis that *m*/z 241.2 represents

18, the diagnostic isotopic pattern of *m*/z 543.2 suggested the presence of a Zn atom (**Figure 3.8**), which would be consistent with a Zn(II)[**18** - H^+]$_2$ metal complex. Such a complex could form through tautomerization and deprotonation of two equivalents of **18** to the corresponding hydroxamate followed by coordination to a Zn^{2+} ion (**19**, **Figure 3.8**). A close analog of **19** derived from aspergillic acid (**20**) has previously been reported from *Aspergillus sojae*.[37]

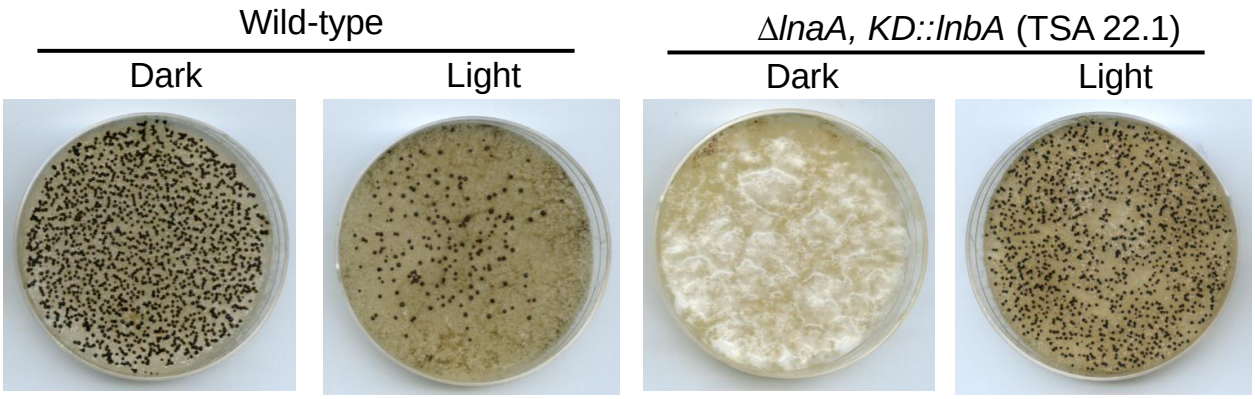


Figure 3.7: *A. flavus* Wild-type and Δ*lnaA, KD::lnbA's* response to light. GMM cultures of wild-type and Δ*lnaA, KD::lnbA* were incubated in either light or dark conditions. Wild-type shows ample sclerotial (small black dots) production in the dark and reduced sclerotial production in the light. Δ*lnaA, KD::lnbA* shows no sclerotial production in the dark and ample sclerotial production in the light.

The initial HPLC/ESI^+-MS-based comparison of wild-type- and TSA 22.1-light extracts were complemented with DANS analysis, demonstrating that compound **19** is upregulated in TSA 22.1 relative to wild-type, and no other metabolites appeared to be differential between these strains (Appendix C, **Figure C.11**). Additional reports of iron(III)- and copper(II)-containing aspergillic acid complexes suggested that profiling the various wild-type and TSA 22.1 extracts for these chemical species may provide a more complete picture of the chemical variability between the wild-type and TSA 22.1 *A. flavus* strains and light/dark conditions.[38,39] Further, because the *A. flavus* cultures were grown on chemically defined GMM media (metal ions: Mn^{2+}, Fe^{3+}, Co^{2+}, Cu^{2+}, Zn^{2+},

Mo^{6+}) a short list of possible aspergillic acid or hydroxy aspergillic acid metal complexes could be generated and quickly searched for in wild-type and TSA 22.1 LC/ESI^+-MS data. For this study LC/ESI^+-MS offered a distinct advantage over NMR-based comparative metabolic profiling because many potential metal-containing chemical species, such as Fe^{3+}-containing derivatives, are paramagnetic, resulting in short nuclear relaxation times and as a result poor NMR spectroscopic line-shape. MS isotopic distribution patterns of transition metals are often diagnostic and, along with exact mass MSMS fragmentation patterns, were used to elucidate structures with high confidence (**Figure 3.8**).

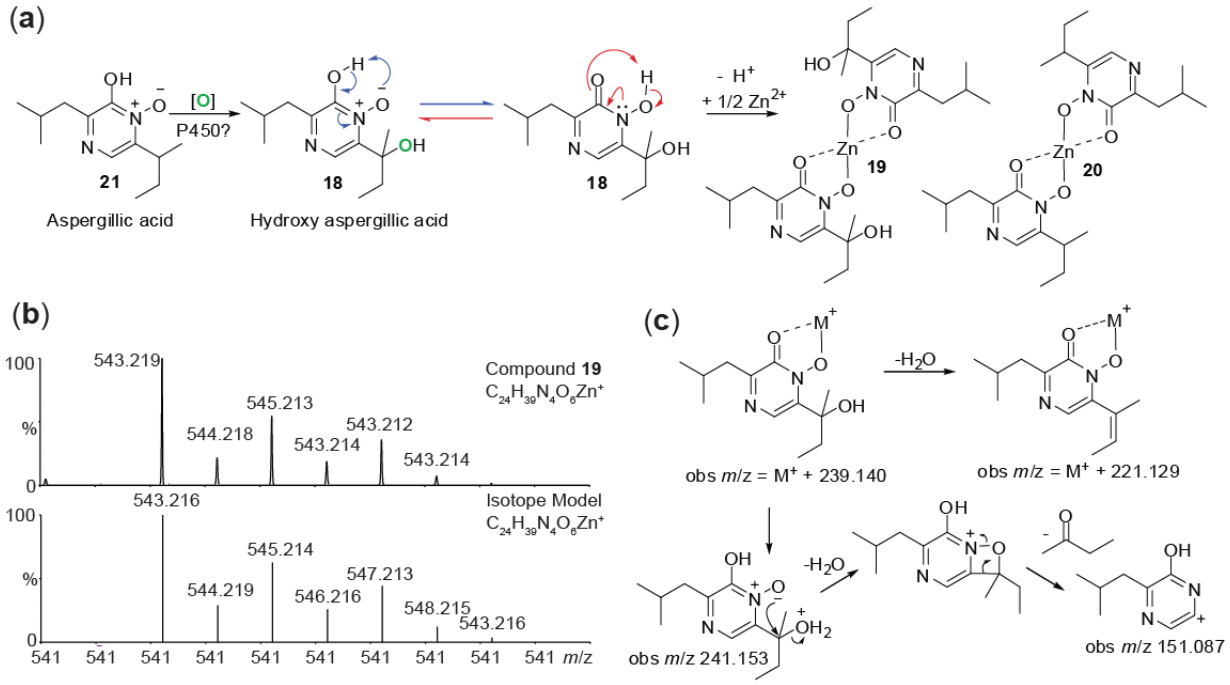


Figure 3.8: Aspergillic acid analogs and MS(MS) analysis. (**a**) Oxidation of aspergillic acid (**21**) to hydroxy aspergillic acid (**18**) followed by tautomerization, deprotonation and coordination to Zn^{2+} forms compound (**19**). (**b**) Example of diagnostic isotopic pattern for compound **19** demonstrating the presence of a zinc ion. (**c**) Fragmentation patterns used to assess the presence of **18**-containing metal complexes. M^+ represents any of the metal centers present in a 2^+ oxidation state.

Qualitative structural elucidation using high resolution UPLC/ESI$^+$-HRMS(MS) profiling of wild-type- and TSA 22.1-light/dark cultures revealed four **18**- or **21**-containing chemical structures, including the previously observed Zn(II)-containing species **19** and **20** as well as two Fe(III)-containing compounds **22** and **23** (**Figure 3.9**, Appendix C, section 13). HPLC/ESI$^+$-MS ion chromatograms acquired for wild-type- and TSA 22.1-light/dark samples were used to compare the relative amounts of **19**, **20**, **22**, and **23** for each sample (**Figure 3.9**).

Compounds **19**, **20**, **22**, and **23** all appear to be upregulated in wild-type-light compared to wild-type-dark. This observation is consistent with previously established models demonstrating that deletion of *laeA* (Δ*laeA*) results in the loss of secondary metabolic production save aspergillic acid,[35] and that the formation of the VelB-VeA-LaeA complex, responsible for coordination of development and secondary metabolisms, is suppressed by light.[36] Interestingly, TSA 22.1 showed **19**, **20**, **22**, and **23** production pattern opposite to wild-type, producing more of these compounds in the dark compared to the light. Lastly, TSA 22.1 showed reduced **20** and **23** production when exposed to the light. These two chemical species both lack hydroxylation of aspergillic acid's sec-butyl moiety, suggesting that, for the TSA 22.1 mutant, the presence of light upregulates aspergillic acid oxidation pathways. As before, the opposite behavior was observed for wild-type.

Further investigations are needed to understand the mechanism by which TSA 22.1 reverses both the phenotypic and metabolic behaviors observed in wild-type sclerotial and aspergillic acid production. The exceptional behavior of the strain TSA 22.1 is even more apparent when compared to TSA 22.6 (also Δ*lnaA, KD::lnbA*) which does not show all of the TSA 22.1 behaviors described here. Taken together these results suggest that the RNAi plasmid used to create TSA 22.1 may have been incorporated into a portion of the genome involved in light-mediated regulation, perhaps by disrupting a yet to be described genetic element involved in controlling aspergillic

acid production. Future efforts to sequence the genomic region adjacent to TSA 22.1's RNAi construct should help generate candidate genes to investigate. Knockout or overexpression of candidate genes may also help to illuminate TSA 22.1's phenoptypic and metabolic mechanisms. Indispensable in this next step will be the LC/ESI$^+$-MS methods developed during this study to characterize and profile resulting metabolic changes.

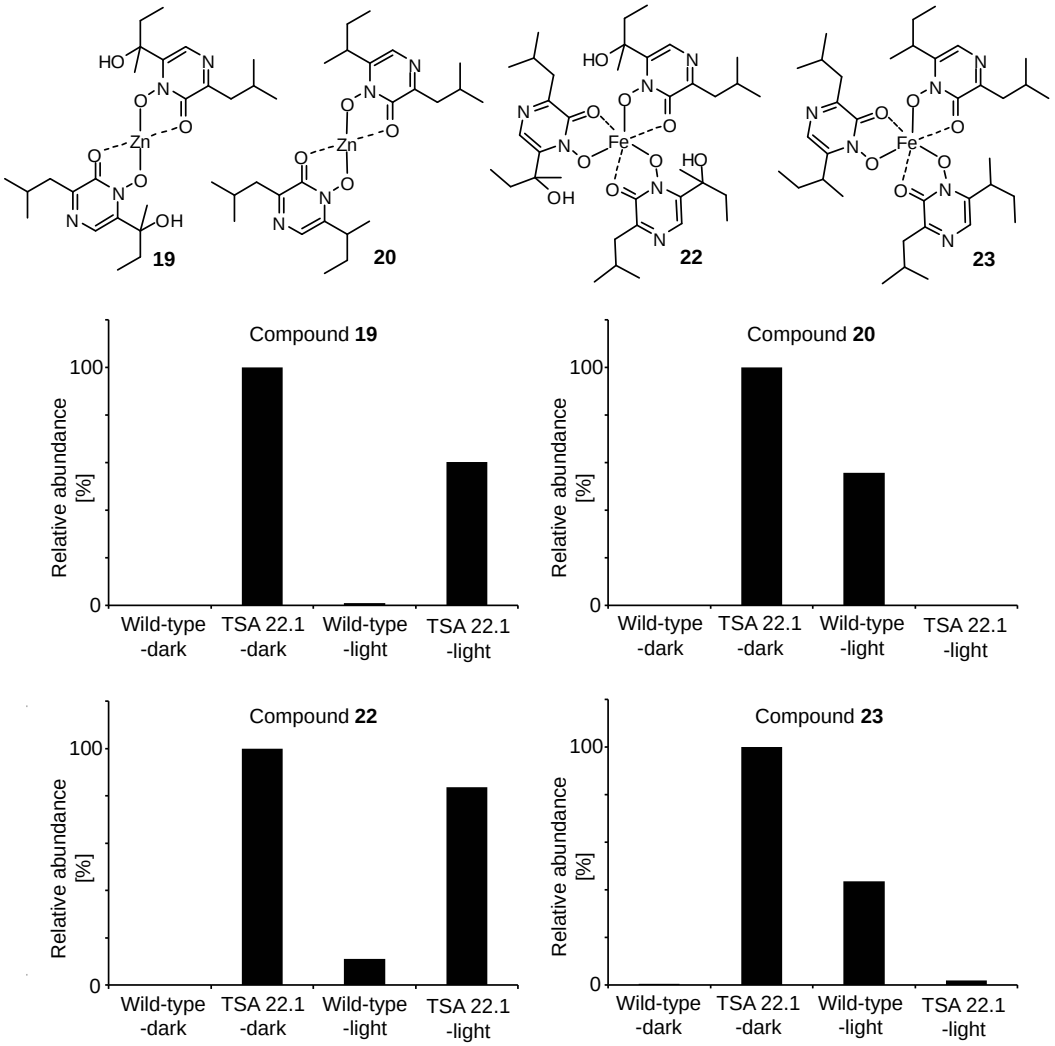


Figure 3.9: Characterized aspergillic acid-derived compounds and wild-type-/TSA 22.1-light/dark sample profiling. Compounds **19**, **20**, **22**, and **23** were characterized *via* UPLC/ESI$^+$-MS(MS) analysis. HPLC/ESI$^+$-MS was then used to profile these chemical species in wild-type- and TSA 22.1-light/dark extracts. Each sample represents 25 pooled culture plates.

REFERENCES

(1) Schneider, P.; Misiek, M.; Hoffmeister, D. *Mol Pharm* **2008**, *5*, 234.

(2) Winter, J. M.; Behnken, S.; Hertweck, C. *Curr Opin Chem Bio* **2011**, *15*, 22.

(3) Brakhage, A. A.; Schroeckh, V. *Fungal Genet Biol* **2011**, *48*, 15.

(4) Georgianna, D. R.; Fedorova, N. D.; Burroughs, J. L.; Dolezal, A. L.; Bok, J. W.; Horowitz-Brown, S.; Woloshuk, C. P.; Yu, J.; Keller, N. P.; Payne, G. A. *Mol Plant Pathol* **2010**, *11*, 213.

(5) Challis, G. L. *J Med Chem* **2008**, *51*, 2618.

(6) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(7) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Bio* **2011**, *15*, 38.

(8) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *PNAS USA* **2009**, *106*, 7708.

(9) Klich, M. A. *Mol Plant Pathol* **2007**, *8*, 713.

(10) Michailides, T.; Thomidis, T. *Plant Pathol* **2007**, *56*, 352.

(11) Hedayati, M. T.; Pasqualotto, A. C.; Warn, P. A.; Bowyer, P.; Denning, D. W. *Microbiology* **2007**, *153*, 1677.

(12) Chakrabarti, A.; Chatterjee, S. S.; Das, A.; Shivaprakash, M. R. *Med Mycol* **2011**, *49 Suppl 1*, S35.

(13) Cleveland, T. E.; Yu, J.; Fedorova, N.; Bhatnagar, D.; Payne, G. A.; Nierman, W. C.; Bennett, J. W. *Trends Biotechnol* **2009**, *27*, 151.

(14) Marui, J.; Yamane, N.; Ohashi-Kunihiro, S.; Ando, T.; Terabayashi, Y.; Sano, M.; Ohashi, S.; Ohshima, E.; Tachibana, K.; Higa, Y.; Nishimura, M.; Koike, H.; Machida, M. *J Biosci Bioeng* **2011**, *112*, 40.

(15) Amaike, S.; Keller, N. P. *Eukaryot Cell* **2009**, *8*, 1051.

(16) Ehmann, D. E.; Gehring, A. M.; Walsh, C. T. *Biochemistry* **1999**, *38*, 6171.

(17) Bergmann, S.; Schumann, J.; Scherlach, K.; Lange, C.; Brakhage, A. A.; Hertweck, C. *Nat Chem Biol* **2007**, *3*, 213.

(18) Yin, W.; Keller, N. P. *J Microbiol* **2011**, *49*, 329.

(19) Kennedy, J.; Turner, G. *Mole Gen Genet : MGG* **1996**, *253*, 189.

(20) Brakhage, A. A. *Microbiol Mol Biol Rev* **1998**, *62*, 547.

(21) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Biol* **2010**, *15*, 38.

(22) Powell, E.; Huang, S. X.; Xu, Y.; Rajski, S. R.; Wang, Y.; Peters, N.; Guo, S.; Xu, H. E.; Hoffmann, F. M.; Shen, B.; Xu, W. *Biochem Pharm.* **2010**, *80*, 1221.

(23) Fujisawa, T.; Igeta, K.; Odake, S.; Morita, Y.; Yasuda, J.; Morikawa, T. *Bioorg. Med Chem* **2002**, *10*, 2569.

(24) Zhang, J. H.; Peng, Q. R.; Zhang, S. X.; Li, Y. B.; Li, S. Q.; Gao, H. X.; Zhou, Z. Q. *J Mol Struct* **2011**, *987*, 34.

(25) Boger, D. L.; Zhou, J. C. *J Am Chem Soc* **1993**, *115*, 11426.

(26) Ayer, W. A.; Singer, P. P. *Phytochemistry* **1980**, *19*, 2717.

(27) Gloer, J. B.; TePaske, M. R.; Sima, J. S. *Journal of Organic Chemistry* **1988**, *53*, 5457.

(28) Schneider, P.; Weber, M.; Rosenberger, K.; Hoffmeister, D. *Chemistry & Biology* **2007**, *14*, 635.

(29) Pospiech, A.; Cluzel, B.; Bietenhader, J.; Schupp, T. *Microbiol-Uk* **1995**, *141*, 1793.

(30) Kessler, N.; Schuhmann, H.; Morneweg, S.; Linne, U.; Marahiel, M. A. *J Bio Chem* **2004**, *279*, 7413.

(31) Kavanagh, K.; Jornvall, H.; Persson, B.; Oppermann, U. *Cell Mol Life Sci* **2008**, *65*, 3895.

(32) Eastwood, D. C.; *et al. Science* **2011**, *333*, 762.

(33) Olson, A.; *et al. New Phytol* **2012**, *194*, 1001.

(34) Bushley, K. E.; Turgeon, B. G. *Bmc Evol Biol* **2010**, *10*.

(35) Kale, S. P.; Milde, L.; Trapp, M. K.; Frisvad, J. C.; Keller, N. P.; Bok, J. W. *Fungal Genet Biol* **2008**, *45*, 1422.

(36) Bayram, O. S.; Bayram, O.; Valerius, O.; Park, H. S.; Irniger, S.; Gerke, J.; Ni, M.; Han, K. H.; Yu, J. H.; Braus, G. H. *Plos Genet* **2010**, *6*.

(37) Tamotsu Yokotsuka, Y. A., and Masaoki Sasaki *J. Agric. Chem. Soc. Japan* **1968**, *42*, 346.

(38) Nurchi, V. M.; Crisponi, G.; Pivetta, T.; Tramontano, E.; Marincola, F. C.; Lachowicz, J. I. *Polyhedron* **2009**, *28*, 763.

(39) Zhu, F.; Wu, J. S.; Chen, G. Y.; Lu, W. H.; Pan, J. H. *Nat Prod Commun* **2011**, *6*, 1137.

CHAPTER 4

CONCLUSIONS AND OUTLOOK

DANS- and LC-MS-Based Comparative Metabolomics in Microorganisms: One of most striking results from the comparative metabolomics *A. fumigatus* and *S. clavuligerus* studies, Chapters 1 and 2 respectively, is the discovery of novel endogenous secondary metabolites from wild-type organisms and secondary metabolic pathways that are relatively well-studied.[1,2] These observations suggest that our current knowledge of metabolites associated with many metabolic pathways may be substantially incomplete.[3] Given the important roles that small-molecules and associated catabolic and anabolic pathways play in every kingdom of life, continued development and incorporation of facile tools to study the metabolomes of biological systems will be indispensable in advancing chemical biology.

As demonstrated in this dissertation, comparative metabolomics incorporating DANS- and LC/ESI-MS-based methodology is highly effective for studying the metabolic pathways of microbial organisms. Chapter 3's treatise on *A. flavus's* orphan *lna*/*lnb* gene clusters represents the first example where DANS was implemented in concert with several modulators of gene expression, such as overexpression, knockout, and knockdown mutations, not only to identify pathway metabolites, but also to develop a broader view of the pathway's role in *A. flavus's* biology. Specifically, perturbations to individual genes in the *lna*/*lnb* pathway at a genetic level could be detected and contextualized using the small-molecule structural information obtained *via* DANS and/or LC/ESI-MS. The results of the *A. flavus* study have established an analytical and genetic platform for the systematic examination of all the small-molecule biosynthetic pathways of *A. flavus*, potentially extending into other microbial and perhaps more complex organisms' biology.

Statistical Treatment of DANS Spectra: Although the DANS methodology discussed in the preceding chapters offers clear advances in metabolite characterizing and profiling technology, it falls short of the ultimate objective of metabolomics, to know the complete content and dynamic ranges of all metabolites in an organism's metabolome ("full characterization").[4] Full characterization is an analytical goal that will always be limited by the sensitivity of the instrumentation used; in the case of NMR and MS roughly μM[5] and nM (*A. flavus* study) concentrations of analytes, respectively. Many metabolomic techniques circumvent full characterization by focusing only on those metabolites in an organism that are have previously been associated with a specific biological readout, for example monitoring gliotoxin levels in *A. fumigatus*-infected mice.[6] DANS also reduces the necessity of full characterization by examining only those metabolites that are differential between comparative sets of biological samples. However, given the highly integrated nature of may biological pathways, only examining metabolites whose production is entirely dependent on a specific pathway may bias one's view of a system's response to specific perturbations (gene distribution, environmental, chemical) and belie other subtle biologically-significant changes.

In pursuit of capturing the most information from a DANS-based comparative metabolomics approaches, one could envision a statistical treatment of 2D NMR spectra to be compared, such that resulting comparative data sets would assign a statistical significance to an observed spectroscopic change. Future efforts with DANS technology will greatly benefit from the incorporation of statistical tests such as principal component analysis, tools that will enable a more quantitative overview of small-molecule metabolism, including subtle metabolic changes.[4,7,8] This work is currently being undertaking by the Schroeder Laboratory, and is called mvaDANS.

Like DNAS, mvaDANS relies on 2D NMR spectra, using comparison of replicates and reference spectra (example, 3 wild-type spectra versus 3 knockout spectra) to quantify and assign statistical significance to changes observed between sets of

dqfCOSY spectra. LC/ESI-MS can also be integrated into the mvaDANS-based approaches, providing analytically orthogonal data sets, also amenable to multivariate analysis.[9]

Although the mvaDANS and LC/ESI-MS statistical approaches offer methodology to more completely capture information from metabolomics studies, they too fall short of the general “full characterization” metabolomics principle discussed above. However, perhaps full characterization becomes less important as one obtains a broader picture of how a biological system responds to a given perturbation. To this end, future efforts combining metabolomic (NMR and MS), proteomic, genomic, and other omic data sets may greatly accelerate our ability to study the chemistry of complex biological systems.

REFERENCES

(1) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(2) Li, B.; Forseth, R. R.; Bowers, A. A.; Schroeder, F. C.; Walsh, C. T. *Chembiochem* **2012**, *13*, 2521.

(3) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Biol* **2011**, *15*, 38.

(4) Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054.

(5) Gokay, O.; Albert, K. *Analytical and bioanalytical chemistry* **2012**, *402*, 647.

(6) Bok, J. W.; Chung, D.; Balajee, S. A.; Marr, K. A.; Andes, D.; Nielsen, K. F.; Frisvad, J. C.; Kirby, K. A.; Keller, N. P. *Infect Immun* **2006**, *74*, 6761.

(7) Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc Chem Res* **2012**, *45*, 288.

(8) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal Chem* **2005**, *77*, 1282.

(9) Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. *Molecular bioSystems* **2012**, *8*, 470.

Appendix A

IDENTIFICATION OF CRYPTIC PRODUCTS OF THE GLIOTOXIN GENE CLUSTER USING NMR-BASED COMPARATIVE METABOLOMICS AND A MODEL FOR GLIOTOXIN BIOSYNTHESIS

1. ***General Analytical Methods and Equipment:*** **(A) NMR spectroscopy**, NMR spectroscopic instrumentation: a Varian INOVA 600 MHz NMR spectrometer (600.0 MHz ^{1}H reference frequency, 150.6 MHz for ^{13}C) equipped with an HCN indirect-detection probe, a Varian 900 MHz NMR spectrometer (899.9 MHz ^{1}H reference frequency, 226.7 MHz for ^{13}C) equipped with a 5 mm ^{1}H (^{13}C/^{15}N) Cryogenic Probe, and a Bruker AVANCE 600 MHz NMR spectrometer (600.7 MHz ^{1}H reference frequency, 149.9 MHz for ^{13}C) equipped with a 5 mm TXI probe. Non-gradient phase-cycled dqfCOSY spectra were acquired using the following parameters: 0.600 s acquisition time, 500-900 complex increments, 8-32 scans per increment. ROESY spectra were acquired using the following parameters: 0.25 s acquisition time, 0.2-0.4 s mixing time, 200-400 increments, 16-32 scans per increment. Gradient and non-gradient HSQC[AD], HMQC, and HMBC[AD] spectra were acquired with these parameters: 0.25 s acquisition time, 300-600 increments, 4-32 scans per increment. ^{1}H,^{13}C-HMBC spectra were optimized for $J_{H,C}$ = 6 Hz. ^{1}H,^{15}N-gHMBCAD spectra were optimized for $J_{H,15N}$ = 8 Hz. Susceptibility-matched NMR tubes (Shigemi) were used for sample amounts smaller than 2 mg. NMR spectra were processed using Varian VNMR and MestreLabs MestReC and MNOVA software packages.

(B) Mass spectrometry, High-resolution mass spectrometry was performed on a LTQ Orbitrap Velos (Thermo Scientific) or a SYNAPT HDMS: (Q-IMS-TOF). HPLC system equipped with a diode array detector and connected to a Quattro II spectrometer (Micromass/Waters) operated in positive-ion electrospray ionization (ESI^{+}) mode was

used for HPLC/ESI$^+$-MS analysis. Data acquisition and processing for the HPLC/ESI$^+$-MS was controlled by Waters MassLynx software.

(C) Chromatography, Flash chromatography was performed using a Teledyne ISCO CombiFlash system. For semi-preparative HPLC Supelco Discovery HS C18 column (25 cm x 10 mm, 5 µm particle diameter). An Agilent Zorbax Eclipse XDB-C8 column (4.6 x 150 mm, 5 µm particle diameter) was used in HPLC/ESI$^+$-MS analysis.

2. Fungal Strains and Experimental Growth Conditions: *A. fumigatus* strains: AF293 (wild-type), TDWC5.6 (Δ*gliZ*:: *A. parasiticus pyrG*; *pyrG1*), ARC2 (Δ*gliP::A*. *parasiticus pyrG pyrG1)* donated by William J. Steinbach; Duke University, *Δ*gliI (*gliI:A. parasiticus pyrG; pyrG1*). Strains were maintained as glycerol stocks and were cultured at 25°C on glucose minimal medium (GMM).[1]

For preparation of spore suspensions, two week old cultures of wild-type, OE, Δ*gliZ*, Δ*gliP*, and Δ*gliI*, grown on GMM plates, were flooded with 1 mL 0.01% Tween 20. 100 µL of the fluid from the flooded plates was diluted with 0.8 mL 0.01% Tween 20 to make the spore suspensions for inoculation. 7 µL of the spore suspension was point inoculated onto the middle of GMM plates. For initial and replicate studies of wild-type, Δ*gliZ*, and Δ*gliI* twenty plates of each strain were cultured. For DANS comparison studies of OE to wild-type, and Δ*gliZ*, three GMM plates of each strain were grown. Cultures were grown for two weeks at 25.0 °C in the dark.

For DANS comparison of Δ*gliP* and wild-type, an additional 2 mL of sterile filtered (0.2 µm, pore size, Nalgene sterile filter) deionized water was pipetted onto the GMM plates prior to inoculation. For these studies sets of three to five GMM plates of each Δ*gliP* and wild-type were set up. Plates were point inoculated and incubated as described in the previous paragraph.

3. Construction of gliI Knock Out Mutants (Δglil): To delete *gliI* in *A. fumigatus*, the plasmid pDWC8.5 was designed as follows. A PCR amplicon of 898 bp of the *gliI* upstream sequence was digested using *SalI* and *EcoRI*, and ligated into *SalI-EcoRI* digested pJW24 vector carrying *A. parasiticus pyrG*. The resulting vector was digested using *BamHI* and *NotI* and ligated with a *BamHI-NotI* digested PCR amplicon of 889 bp of the *gliI* downstream sequence. This final vector was designated as pDWC8.5. To generate the Δ*gliI* strain (TDWC7.49 and TDWC7.54), DNA of pDWC8.5 was transformed into *A. fumigatus* wild-type AF293.1. Homologous recombination of the *gliI* knock-out construct was verified by PCR and Southern blot analysis (**Figure A.8**).

4. Assesment of gliZ and gliP Expression in Δglil Mutant: Northern analysis was used to assess gene expression in *A. fumigatus* strains. Total RNA was extracted from lyophilized mycelia of *A. fumigatus* using TRIzol (Invitrogen) and separated on a denaturing formaldehyde/agarose gel. RNA was transferred from the gel to a nylon membrane (Hybond^{TM}-N^{+}, Amersham Pharmacia Biotech) by capillary blotting with 10 x SSC. RNA was irreversibly bound to the membrane using a UV cross linker at 1200 J^{3}cm^{-2}. Blots of *A. fumigatus* RNA were hybridized with the following fragments, which had been amplified from *A. fumigatus* genomic DNA and labeled with [^{32}P] αdCTP: a 653 bp fragment of *gliI* (primers 5'-TGTTGATCGAGACGCCGTTCTG-3' and 5'-CAGAGCGGCTCGATTCTGGTG-3'), a 1211 bp fragment of *gliZ* (primers 5'-AAGGGCCGGTAGTCTACCTCTTC-3' and 5'-CGATCTGGTAGCTGCCCAGCTGGAAG-3') and a 173 bp fragment of *gliP* (primers 5'-AAACCCCTGTGAATGCAGAC-3' and 5'-CCCCTTGAGATGAAAGGTGA-3') (**Figure A.8**).

5. Metabolome Extraction and Initial Fractionation: Fungal cultures including fungal tissue and agar media were cut into small ~1 x 1 cm cubes and placed into individual 1 L round bottom flasks, frozen using a dry ice acetone bath, and lyophilized. Two

different protocols were used for extraction of fungal cultures: (1) lyophilized wild-type and Δ*gliZ* cultures were extracted with ethyl acetate and a mixture of 5% ethyl acetate, 15% water, and 80% acetonitrile; (2) for additional analysis of wild-type and Δ*gliZ* cultures and for analyses of the Δ*gliP* and Δ*gliI* cultures 100% ethyl acetate followed by 100% acetonitrile were used to extract the lyophilized fungal cultures. All extracts were filtered over Celite and a glass frit. Subsequently, the extracts were adsorbed onto Celite (~1 g of Celite for 100 mg of dried extract) and fractionated on a 100 g RediSep Reverse-Phase (RP) C18 flash chromatography column using a water-acetonitrile solvent gradient, starting with 20 min of 5% acetonitrile followed by a linear increase to 100% acetonitrile over 40 min, followed by 10 min of 100% acetonitrile. Fractions were combined into three pools; pool 1, 0.5% acetonitrile isocratic elution, fractions 21-44; pool 2, 0.5-55% acetonitrile, fractions 45-101; pool 3, 55-100% acetonitrile, fractions 102-140. Pools were evaporated to dryness and suspended in ~0.2 mL of acetonitrile-*d*3 (in the case of pool 1, ~0.2 mL 1:2 D_2O:acetonitrile-*d*3). The suspension was centrifuged to remove insoluble materials; this was repeated two additional times, and the supernatant was subjected to NMR spectroscopic analysis.

6. Differential Analysis by 2D-NMR Spectroscopy (adapted from reference[2]): High-resolution dqfCOSY spectra were acquired using the following parameters: acquisition time 0.6 s; 500 increments (ni). Phase cycling was used for coherence selection, and MestReC was used to process the resulting data, zero-filling the spectra to 8096 complex data points in the directly detected dimension (F2) and 4096 data points in the indirectly detected dimension (F1). Bitmaps derived from absolute-value processed dqfCOSY spectra were then imported into Adobe Photoshop CS3 and overlayed as described in reference.[3]

Figure A.1: GliP domain architecture. GliP includes two adenylation (A), three thiolation (T), and two condensation domains.[4]

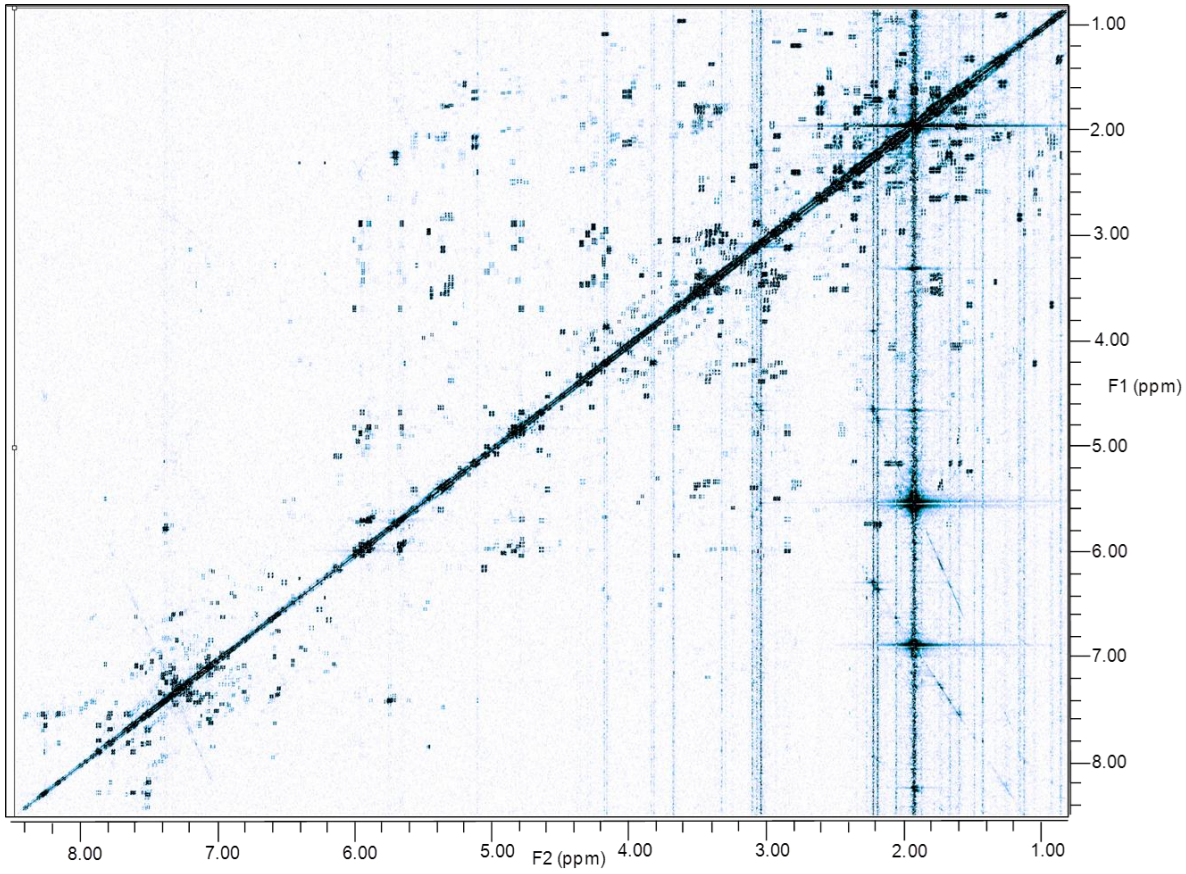


Figure A.2: Magnitude-mode processed dqfCOSY spectrum of *A. fumigatus* wild-type (600 MHz, solvent acetonitrile-d_3).

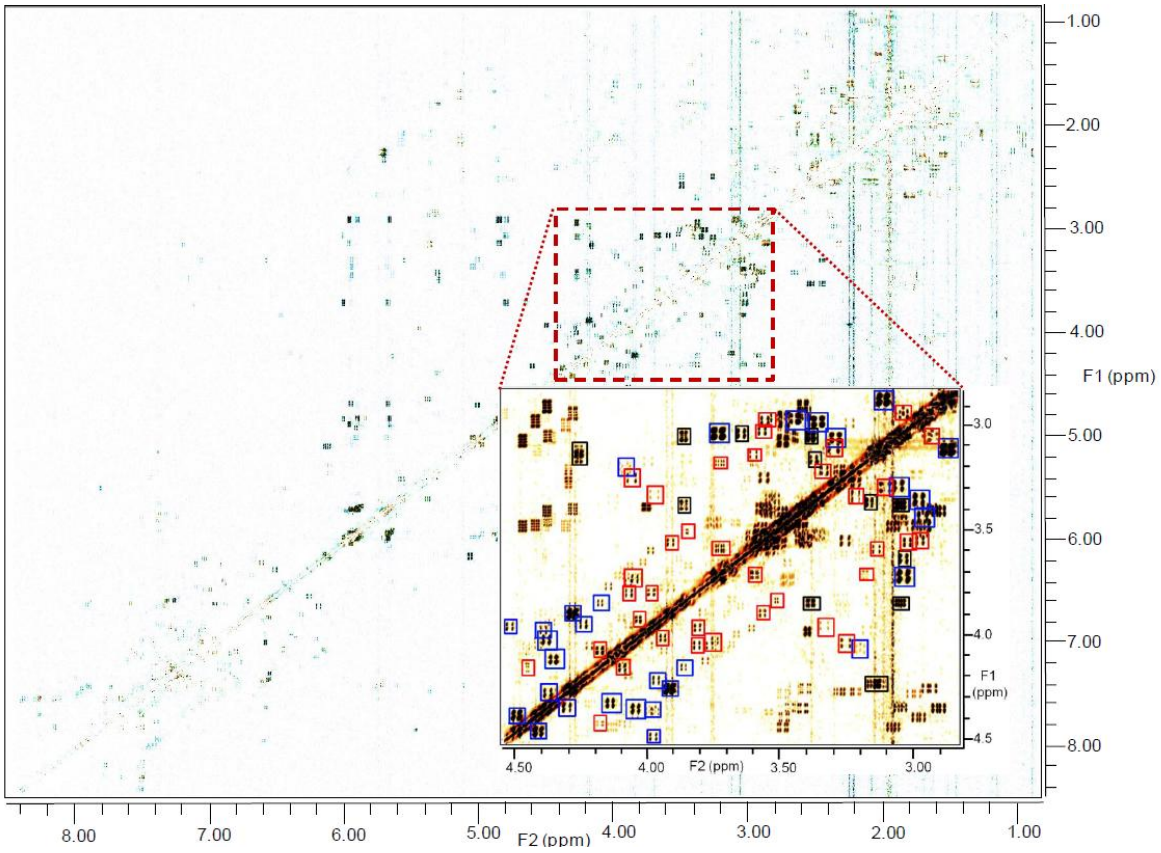


Figure A.3: DANS overlay of wild-type pool 2 and *ΔgliZ* pool 2, based on magnitude-mode processed dqfCOSY spectra (600 MHz, acetonitrile-*d*3). Insert: 4.55-2.75 ppm section enhanced to reveal minor components. *gliZ*-dependent signals detected *via* DANS are boxed, representing known gliotoxin derivatives (blue), the known **10** and **11** (black), and novel compounds **12**-**17** (red).

7. Characterization of Compounds 10-14a/b, 16, and 17: The wild-type-derived pools used for DANS analysis described above were fractionated using a Supelco Discovery HS C18 column (25 cm x 10 mm, 5 µm particle diameter) and a water-acetonitrile solvent gradient, starting with 5 min at 5% acetonitrile, followed by a linear increase to 23% acetonitrile over 1 min, and a more gradual increase reaching 50% acetonitrile at 49 min. A total of 18 fractions were collected. Fractions 1-18 were evaporated to dryness *in vacuo* and dissolved in acetonitrile-d_3, followed by ^{1}H-NMR spectroscopic analysis of each fraction. Comparison with data from DANS analyses indicated that

fraction 1 contained nearly pure compound **10**, and fraction 5 contained nearly pure compound **11**. Comparison with the DANS data further indicated that compounds **12**, **14b**, and **16** were present in fraction 2, and compounds **13**, **14a**, and **17** in fraction 4.

Fractions 2 and 4 were further purified using the same column and solvent system described above. Fraction 2 was separated to obtain samples of **12**, **14b**, and **16** using a solvent gradient starting at 5% acetonitrile for 30 min, followed by a linear increase to 20 % acetonitrile over 30 min, which was followed by 10 min at 20% acetonitrile. Fraction 4 was fractionated further to obtain samples of **13**, **14a**, and **17** using a solvent gradient of 30 min at 5% acetonitrile with a linear increase to 20 % over 5 min followed by 12 min 20% acetonitrile. In each case collected fractions were evaporated to dryness *in vacuo* and dissolved in acetonitrile-d_3, followed by NMR spectroscopic analysis of each fraction.

8. Purification of Compound 4: 20 wild-type cultures grown on GMM plates, as described above, were extracted using a mixture of 5% ethyl acetate, 15% water, and 80% acetonitrile (section 2). These extracts were initially used to determine an appropriate enrichment protocol for compound **4**. Wild-type extracts were dissolved in ~50 % acetonitrile/water and absorbed onto Celite. Solvent was removed *in vacuo*. The sample, absorbed onto Celite, was subjected to flash chromatography using a CombiFlash chromatography system equipped with a 100 g RediSep Gold RP-18 flash chromatography column using acetonitrile and 0.1% acetic acid in water as solvents, starting with 3.5 min at 1% acetonitrile, followed by a linear increase to 10% acetonitrile over 16 min, which was followed by an increase to 100% acetonitrile over 3.5 min, holding at 100% acetonitrile for 3.5 min. Fractions were evaporated to dryness, and ^{1}H NMR spectra of individual fractions were acquired in methanol-d_4. Signals diagnostic for compound **4**, including signals representing olefinic, methylene, and methyl sulfanyl protons, indicated that **4** elutes at ~9% acetonitrile under these conditions. Fractions

containing **4** were pooled together. For purification of a larger amount of compound **4**, this procedure described above was repeated twice more using extracts from two 90-plate-batches of wild-type. Fractions containing **4** were pooled to give a single sample of enriched **4** which was subjected further 2D NMR spectroscopic and HR-MS analysis.

9. Isolation of Compounds 15a/b: The remaining extract from the 200 wild-type cultures used for isolation of compound **4** was adsorbed onto Celite and subjected to flash chromatography using a CombiFlash chromatography system equipped with a 100 g RediSep Gold RP-18 column. The water-acetonitrile solvent gradient used is described in section 5 of Appendix A. Fractions eluting at 10-50% acetonitrile were pooled and evaporated *in vacuo*. The residue was dissolved in acetonitrile, filtered over acetonitrile-washed cotton, and subjected to preparative HPLC using instrumental setup and solvent systems described in section 6 of Appendix A. A water-acetonitrile solvent gradient was used that started at 5% acetonitrile for 30 min, followed by a linear increase to 20% acetonitrile over 30 min. (Compounds **15a/b** were found to elute at 13-14% acetonitrile under these conditions) Solvent was removed from the HPLC fractions *in vacuo* and samples were redissolved in acetonitrile-d_3, and ^{1}H NMR spectra were acquired for each HPLC fraction. Diagnostic signals present in the DANS spectrum of pool #2 indicated the presence of **15a/b**.

10. High Resolution Mass Spectrometric Analysis for Compounds 4, 5-7, 9, 12, and 13-17: Enriched samples of compound **4**, **5**, **7**, **9**, and **15a/b** were diluted to 0.01 mg/mL by dissolving the samples in either 20% acetonitrile and 80% of 0.1% formic acid in water, **4**; or 50% acetonitrile and 50% of 0.1% formic acid in water, **5**-**7**, **9** and **15a/b**. These sample were infused into a LTQ Orbitrap Velos mass spectrometer running in ESI$^+$ ionization mode with a mass range at or within 1-2000 *m*/z.

Compound	HR-ESI$^+$MS obseved (m/z)	Ion	Calculated for Ion Formula	Calculated m/z
4	351.0434	$[M+Na]^+$	$C_{13}H_{16}N_2NaO_4S_2$	351.0438
5	325.0321	$[M+H]^+$	$C_{13}H_{13}N_2O_4S_2$	325.0311
6	357.0039	$[M+H]^+$	$C_{13}H_{13}N_2O_4S_3$	357.0032
7	388.9760	$[M+H]^+$	$C_{13}H_{13}N_2O_4S_4$	388.9753
9	377.0601	$[M+Na]^+$	$C_{15}H_{18}N_2NaO_4S_2$	377.0600
15a/b	317.0564	$[M+Na]^+$	$C_{13}H_{14}N_2NaO_4S$	317.0566

Samples of compounds **12**, **13**, **14a/b**, **16**, and **17** were diluted to a concentration 0.01 mg/mL in acetonitrile. An aliquot of the diluted sample was acidified using formic acid (final concentration 0.1%). Samples **12**, **13**, **14a/b**, **16**, and **17** were infused into a SYNAPT HDMS: (Q-IMS-TOF). Mass spectra were acquired using ESI^+ ionization with a mass range of 1-1000 *m*/z.

Compound	HR-ESI$^+$MS obseved (m/z)	Ion	Calculated for Ion Formula	Calculated m/z
12	281.107	$[M+H]^+$	$C_{13}H_{17}N_2O_3S$	281.095
13	365.055	$[M+Na]^+$	$C_{14}H_{18}N_2NaO_4S_2$	365.060
14a/b	333.106	$[M+Na]^+$	$C_{14}H_{18}N_2NaO_4S$	333.088
16	303.087	$[M+Na]^+$	$C_{13}H_{16}N_2NaO_5$	303.095
17	261.079	$[M+H]^+$	$C_{13}H_{13}N_2O_4$	261.087

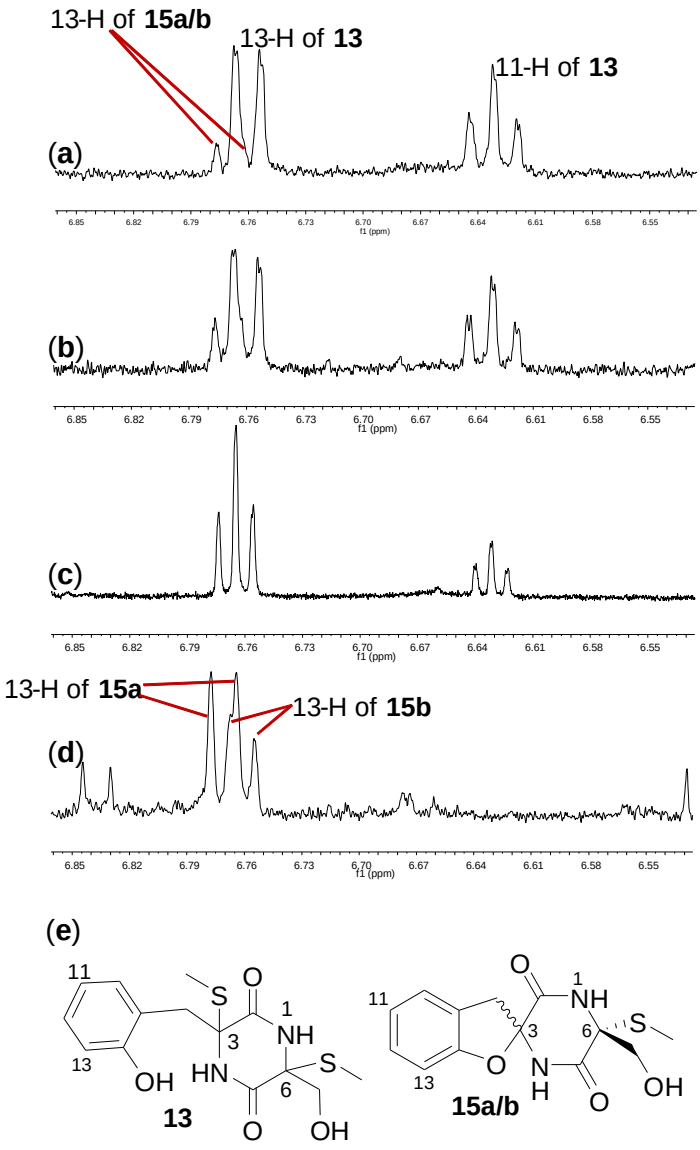


Figure A.4: Formation of a mixture of diastereomers of spirocyclic 15 from diketopiperazine 13. (**a**) 6.5-6.9 ppm section of the ^{1}H NMR spectrum of a sample of diketopiperazine **13**, showing signals for protons 11-H and 13-H of **13** in addition to a small signal for 13-H of **15a** and **15b** (600 MHz, DMSO-d_6). (**b**) Same sample after several days at room temp., showing increasing amounts of **15a/b** (600 MHz, DMSO-d_6). (**c**) After two weeks at room temp., most of **13** has converted to **15a/b** (900 MHz, DMSO-d_6). (**d**) Sample of a mixture of **15a** and **15b** (600 MHz, DMSO-d_6), containing small amounts of impurities (signals at 6.84 and 6.53 ppm). (**e**) Compounds **13** and **15a/b**.

11. HPLC-MS-based Comparison of Wild-Type and Mutant Extracts: Unfractionated wild-type and mutant extracts prepared as described in section 2 were dissolved in acetonitrile (~1 mL solvent/10 mg sample). Post-HPLC (using the RP-8 column) and UV spectroscopic analysis (200-400 nm), samples were infused into a Quattro II ESI MS running in ESI^+ mode. Enriched samples of NMR- and MS-characterized *gliZ*-dependent metabolites served as elution time and ionization standards. Testing for the presence of compounds **1**-**3**, **8**, **10**-**12**, **14a/b**, and **17** in wild-type and mutant extracts a water/acetonitrile gradient was used starting with 5 min at 5% acetonitrile followed by a linear increase to 20% acetonitrile over 5 min, this was followed by a linear increase to 45% acetonitrile over 110 min (for chromatograms, see **Figure A.5** and **Figure A.6**. For testing the presence or absence of compounds **13** and **16** in wild-type and Δ*gliZ* extracts a different water/acetonitrile gradient was used starting with 5 min at 5% acetonitrile followed by a linear increase to 25% acetonitrile over 55 min.

12. Relative Abundance of gliZ-Dependent Compounds: Using NMR solvent signals as internal standards, relative abundances of compounds **1**-**17** were estimated based on integration of their ^{1}H NMR spectroscopic signals in partially purified samples derived from a set of 20 wild-type plates. Abundance of *gliZ*-dependent metabolites relative to that of gliotoxin (**1**): compound **4**, 30:1; compound **10**, 2:1; compound **12**, 4:1; compound **11**, 8:1; compound **13**, 8:1; compound **14a**, 20:1; compound **14b**, 60:1; compound **16**, 50:1; compound **17**, 10:1, and compounds **2**, **3** and **8** ~1:1**.** Abundance of compounds **4** and **16** are likely underestimated significantly because of partial decomposition of these compounds during purification.

13. DANS of ΔgliP and Wild-Type: GMM plates were prepared as described in section 2, except that an additional 200 µL of sterile filtered (0.2 µm, pore size, Nalgene sterile filter) deionized water was pipetted onto the GMM plates prior to inoculation. For these

studies sets of five GMM plates of each Δ*gliP* and wild-type were used. Plates were point inoculated, incubated, and extracted as described in above (sections 2 and 5).

14. ΔgliP Feeding Experiment: In parallel to the wild-type and Δ*gliP* cultures described above, five plates of Δ*gliP* supplied with synthetic compound **10** ("Δ*gliP*+**10**") were grown. For these cultures, 13 mg of **10**, purchased from Sigma-Aldrich, was dissolved in 3 mL deionized water and sterile filtered. 200 µL of this solution was pipetted onto each plate. Subsequently, plates were inoculated, incubated, and extracted as described above (Apendix A, sections 2 and 5). DANS comparison as well as HPLC-MS analysis of extracts derived from wild-type (positive control), Δ*gliP* (negative control), and Δ*gliP*+**10** (treatment) revealed that none of the *gliZ*-dependent metabolites were rescued by addition of **10** to Δ*gliP*.

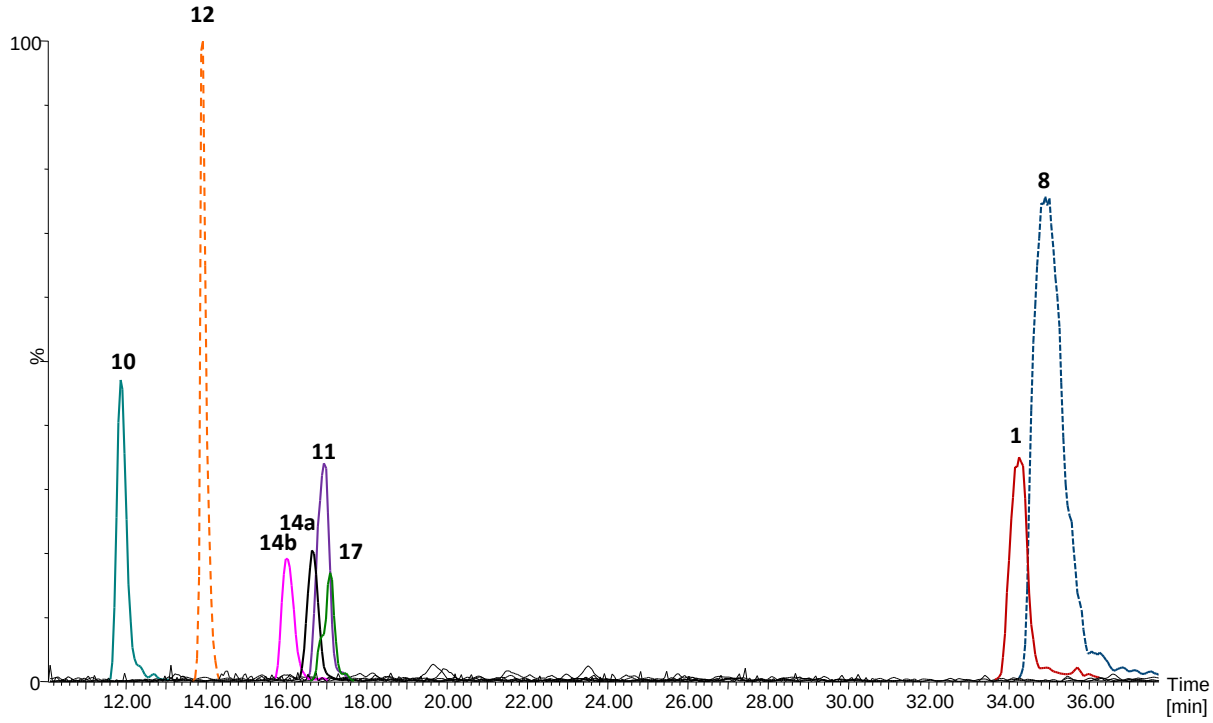


Figure A.5: HPLC/ESI$^+$-MS ion chromatograms for *gliZ*-dependent metabolites 1, 8, 10, 11, 12, 14a, 14b, and 17. Ion chromatograms were obtained *via* HPLC-MS analysis of partially purified samples derived from *A. fumigatus* wild-type cultures. For each compound, a chromatogram representing the most abundantly produced ion is shown: m/z compound **1**, $[M\text{-}S_2+H]^+$ = 263.1; compound **10**, [M+1] = 235.1; compound **12**, $[M\text{-}SMe]^+$ = 233.1; compound **11**, $[M\text{-}SMe]^+$ = 279.1; compound **14a**, $[M\text{-}SMe]^+$ = 263.1; compound **14b**, $[M\text{-}SMe]^+$ = 263.1; compound **17**, $[M+H]^+$ = 261.1; and compound **8**, $[M+H]^+$ = 357.1.

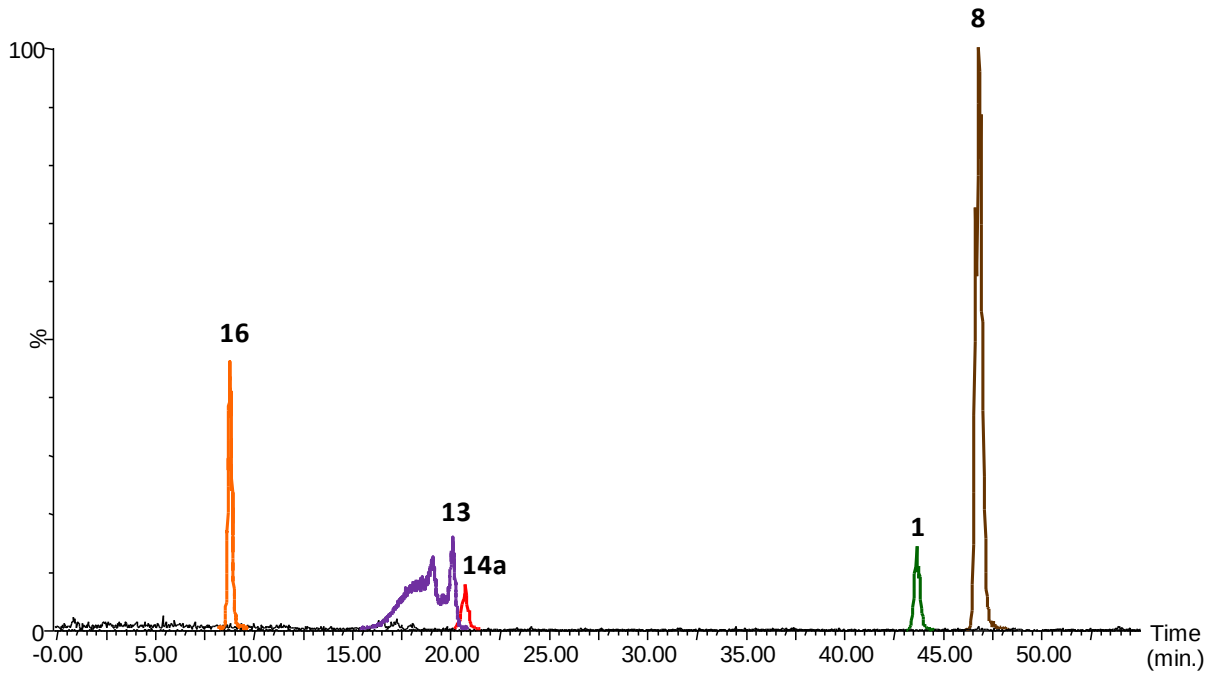


Figure A.6: HPLC/ESI$^+$-MS ion chromatograms for *gliZ*-dependent metabolites 1, 8, 13, 14a, and 16. Chromatograms were obtained *via* HPLC/ESI$^+$-MS analysis of partially purified samples derived from *A. fumigatus* wild-type cultures. Chromatograms for compounds **13** and **16** were obtained using single-ion monitoring MS. For each compound, a chromatogram representing the most abundantly produced ion is shown: *m*/z compound **16**, [M+1] = 281.1; compound **13**, [M-SMe]$^+$ = 295.1; compound **14a**, [M-SMe]$^+$ = 263.1; compound **1**, [M-S_2+H]$^+$ = 263.1; and compound **8**, [M+H]$^+$ = 357.1.

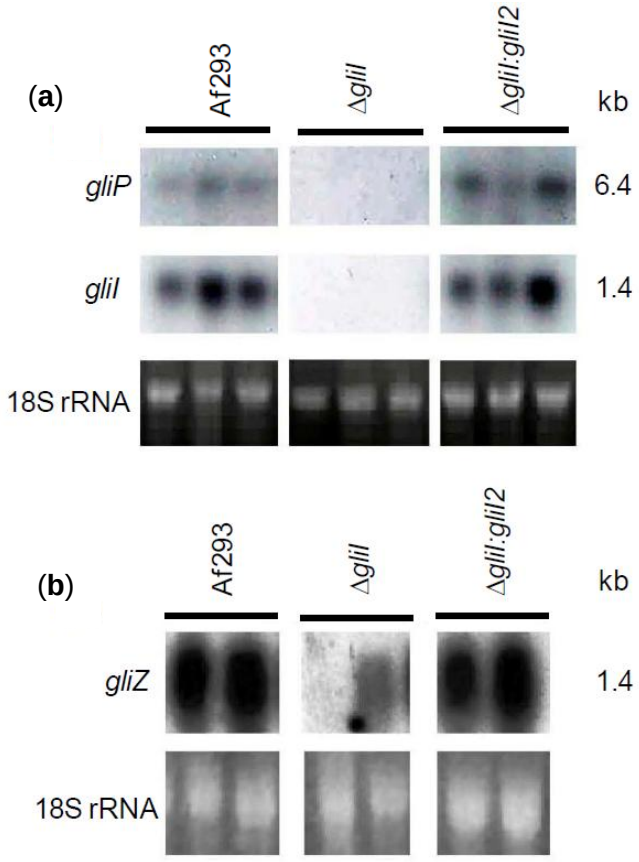


Figure A.7: *gliP*, *gliI*, and *gliZ* expression in the wild-type (Af293) Δ*gliI,* Δ*gliI:gliI2* strains. Total RNA was extracted from shaking cultures grown in Czapek Dox medium at 37 ^{o}C for two days. RNA was probed with a fragment of gliotoxin biosynthetic genes *gliP* (173 bp), *gliI* (653 bp) and *gliZ* (1211 bp)**.** Transcript sizes are marked. Ethidium bromide-stained 18S rRNA is shown as a control for RNA loading. (**a**) Transcripts of *gliP* and *gliI* were undetected in the *ΔgliI* mutant, but present in similar amounts in isolates Af293 and *ΔgliI:gliI*2 (Δ*gliI* complemented with *gliI*). (**b**) Transcripts of *gliZ* were not detected in the *ΔgliI* mutant in one biological replicate and reduced in the second biological replicate, compared to that in Af293. The level of *gliZ* transcription was restored to wild-type level in isolate *ΔgliI:gliI2* in both biological replicates

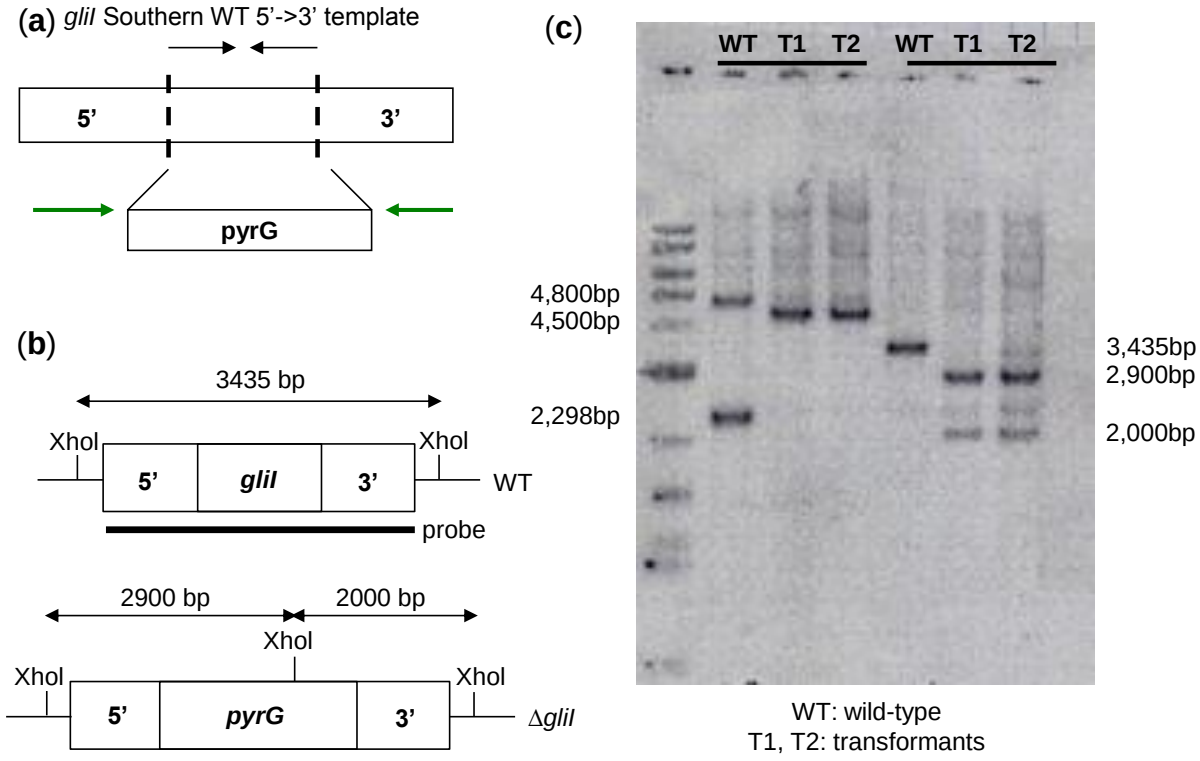
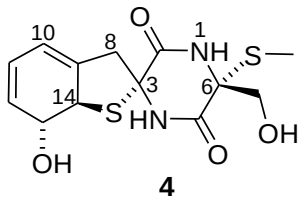


Figure A.8: Δ*gliI* mutants' construction. The construction of Δ*gliI* mutants was confirmed by Southern analysis of genomic DNA from wild-type (WT) and two transformants (T1 and T2), indicating *gliI* replacement by *pyrG* in T1 and T2. (**a**) Wild type and two *gliI* replacement strains were probed with a 3,435 bp fragment of DNA covering approximately 1 kb up and down stream of *gliI* ORF as well as the *gliI* ORF. (**b**) XhoI does not cut within the *gliI* ORF, rather XhoI cuts regions flanking *gliI* generating a single DNA fragment of 3,435 bp. The DNA from transformants in which *pyrG* replaces *gliI* is cut by XhoI into two fragment of 2,900 and 2,000 bp each. (**c**) Southe**r**n blot analysis of wild-type and T1 and T2 after DNA treatment with BamHI and XhoI agree with predicted XhoI treatment DNA restriction products, confirming that *gliI* is present in wild-type and has been replaced by *pyrG* in T1 and T2.

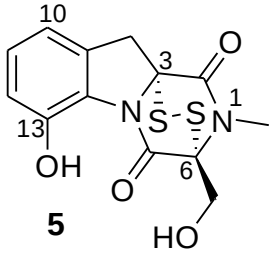
Table A.1: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **4** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via HMQC and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon. ROESY correlations were observed using a mixing time of 275 ms. Abbreviation: wk = weak, but clearly discernible ROESY correlation.



4

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**	**ROESY**
1-NH		1-NH	9.32	3, 5	7-H_a, 6-S-CH_3
2	167.9				
3	69.6				
4-NH		4-NH	9.07	2, 5, 6, 8	8-H_a, 13-H (wk)
5	165.1				
6	65.3				
6-S-CH_3	12.2	6-S-CH_3	2.12		7-H_a 7-H_b, 14-H (wk)
7	63.6	7-H_a	3.57 ($J_{7\text{-Ha},7\text{-Hb}}$ = 11)	5, 6	7-H_b
		7-H_b	3.88	2, 5, 6	
8	44.7	8-H_a	2.89 ($J_{8\text{-Ha},8\text{-Hb}}$ = 16, $J_{8\text{-Ha},10\text{-H}}$ = 2, $J_{8\text{-Ha},12\text{-H}}$ <1, $J_{8\text{-Ha},13\text{-H}}$ < 1)	2, 3, 9, 10, 14	7-H_b, 10-H
		8-H_b	3.40 ($J_{8\text{-Hb},10\text{-H}}$ = 2, $J_{8\text{-Ha},12\text{-H}}$ <1, $J_{8\text{-Hb},13\text{-H}}$ < 1, $J_{8\text{-Hb},14\text{-H}}$ = 2)	2, 3, 9, 10, 15	14-H, 10-H
9	141.1				
10	115.8	10-H	5.77 ($J_{10\text{-H},11\text{-H}}$ = 6.7)	14	11-H
11	123.9	11-H	5.85 ($J_{11\text{-H},12\text{-H}}$ = 10, $J_{11\text{-H},13\text{-H}}$ = 3)	9	12-H
12	133.1	12-H	5.66 ($J_{12\text{-H},13\text{-H}}$ = 3, $J_{12\text{-H},14\text{-H}}$ < 1)	10, 13, 14	13-H
13	73.3	13-H	4.44 ($J_{13\text{-H},14\text{-H}}$ = 15)	9, 14	
14	57.1	14-H	4.21	9, 13	7-H_a, 6-S-CH_3

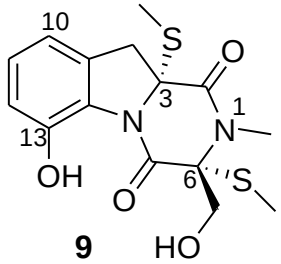
Table A.2: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **5** in acetonitrile-d_3. Chemical shifts were referenced to δ(C<u>H</u>D_2CN) = 1.94 ppm and δ(^{13}<u>C</u>HD_2CN) = 1.3 ppm. ^{13}C chemical shifts were determined via HMQC and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon. Abbreviation: wk = weak, but clearly discernible HMBC.



5

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-N				
1-N-CH_3	27.8	1-N-CH_3	3.20	2, 6
2	165.9			
3	74.6			
4-N				
5	162.7			
6	78.1			
7	60.4	7-H_a	4.35 ($J_{7\text{-Ha},7\text{-Hb}}$ = 13)	5, 6
		7-H_b	4.46	5
7-OH		7-OH		
8	36.9	8-H_a	3.40 ($J_{8\text{-Ha},8\text{Hb}}$ = 19, $J_{8\text{-Ha},10\text{-H}}$ < 1, $J_{8\text{-Ha},12\text{-H}}$ < 1)	2, 3, 8, 9, 10, 11(wk), 14
		8-H_b	4.19 ($J_{8\text{-Hb},10\text{-H}}$ < 1, $J_{8\text{-Hb},12\text{-H}}$ < 1)	2, 3, 8, 9, 10, 11(wk), 13, 14
9	132.0			
10	117.1	10-H	6.88 ($J_{10\text{-H},11\text{-H}}$ = 8)	8, 14
11	129.6	11-H	7.16 ($J_{11\text{-H},12\text{-H}}$ = 7)	9, 13
12	117.9	12-H	6.83	14
13	145.9			
13-OH		13-OH	10.73	11(wk), 12, 13, 14
14	124.9			

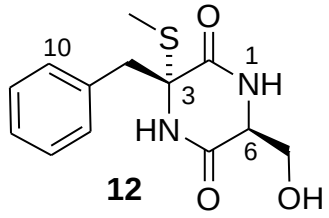
Table A.3: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **9** in acetonitrile-d_3. Chemical shifts were referenced to δ(C<u>H</u>D_2CN) = 1.94 ppm and δ(^{13}<u>C</u>HD_2CN) = 1.3 ppm. ^{13}C chemical shifts were determined via HMQC and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon. Abbreviation: wk = weak, but clearly discernible HMBC.



9

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-N				
1-N-CH_3	29.0	1-N-CH_3	3.12	2, 6
2	165.8			
3	72.1			
3-S-CH_3	14.8	3-S-CH_3	2.21	3
4-N				
5	165.7			
6	73.4			
6-S-CH_3	13.6	6-S-CH_3	2.31	6
7	64.4	7-H_a	3.91 ($J_{7\text{-Ha},7\text{-Hb}}$ = 12, $J_{7\text{-Ha},7\text{-OH}}$ = 2)	2, 6
		7-H_b	4.29 ($J_{7\text{-Hb},7\text{-OH}}$ = 2)	2
7-OH		7-OH	5.74	
8	40.0	8-H_a	3.47 ($J_{8\text{-Ha},8\text{Hb}}$ = 17, $J_{8\text{-Ha},10\text{-H}}$ < 1, $J_{8\text{-Ha},12\text{-H}}$ < 1)	3, 9, 10, 14
		8-H_b	3.55 ($J_{8\text{-Hb},10\text{-H}}$ < 1, $J_{8\text{-Hb},12\text{-H}}$ = < 1)	2, 3, 9, 10, 13(wk), 14
9	132.9			
10	117.4	10-H	6.90 ($J_{10\text{-H},11\text{-H}}$ = 8)	8, 12, 13, 14
11	129.8	11-H	7.19 ($J_{11\text{-H},12\text{-H}}$ = 7)	9, 13
12	118.1	12-H	6.86	10, 13, 14
13	147.3			
13-OH		13-OH	10.27	11, 12, 13, 14
14	128.6			

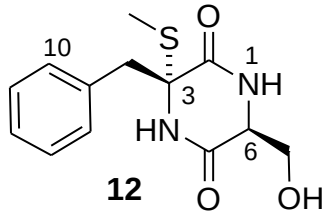
Table A.4: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **12** in acetonitrile-d_3. Chemical shifts were referenced to δ($C\underline{H}D_2CN$) = 1.94 ppm and δ($^{13}\underline{C}HD_2CN$) = 1.3 ppm. ^{13}C chemical shifts were determined via HSQCAD and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



12

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-NH		1-NH	6.40 or 6.85	
2	166.2			
3	69.1			
3-S-CH_3	13.2		2.18	3
4-NH		4-NH	6.40 or 6.85	
5	166.5			
6	57.6	6-H	3.15 ($J_{6\text{-H},7\text{-Ha}}$ = 7, $J_{6\text{-H},7\text{-Hb}}$ = 6)	5
7	63.5	7-H_a	3.53 ($J_{7\text{-Ha},7\text{-Hb}}$ = 12, $J_{7\text{-Ha},7\text{-OH}}$ = 8)	
		7-H_b	3.66	
7-OH		7-OH	3.17	
8	45.6	8-H_a	2.96 ($J_{8\text{-Ha},8\text{-Hb}}$ = 13)	2, 3, 9, 10, 14
		8-H_b	3.51	3, 9, 10, 14
9	135.7			
10, 14	131.5	10-H, 14-H	7.25	8, 10, 12, 14
11, 13	129.1	11-H, 13-H	7.30	9, 11, 13
12	128.6	12-H	7.27	10, 14

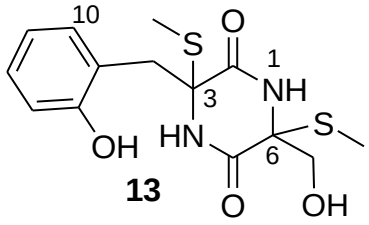
Table A.5: ^{1}H (600 MHz) NMR spectroscopic data for compound **12** in DMSO-d_6. Chemical shifts were referenced to δ(C<u>H</u>D_2SOCD_3) = 2.50 ppm. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum. ROESY correlations were observed using a mixing time of 350 ms. Coupling multiplicities are annotated as: s, singlet; d, doublet; dd, doublet of doublets; m, multiplet.



12

Position	Proton	δ_H (*J* [Hz])	ROESY
1-NH	1-NH	8.04 (s)	6-H, 7-H_a, 7-OH
2			
3			
3-S-CH_3	3-S-CH_3	2.18 (s)	4-H, 8-H_b, 7-OH
4-NH	4-NH	8.80 (s)	8-H_a, aromatic-H (7.20)
5			
6	6-H	3.14 (m)	7-OH
7	7-Ha	3.42 (m)	7-H_b, 7-OH
	7-Hb	3.63 (m)	7-H_a, 7-OH
7-OH	7-OH	4.96 (dd, *J* = 5, *J* = 6) (No signal upon D_2O addition)	
8	8-Ha	2.92 (d, *J* = 13)	8-H_b, aromatic-H (7.24)
	8-Hb	3.41 (d, *J* = 13)	8-H_a, aromatic-H (7.24)
9			
phenyl		7.25 - 7.21	

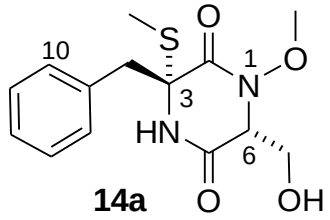
Table A.6: ^{1}H (900 MHz) and ^{13}C (227 MHz) NMR spectroscopic data for compound **13** in DMSO-*d*$_6$. Chemical shifts were referenced to δ(C<u>H</u>D$_2$SOCD$_3$) = 2.50 ppm and δ(^{13}<u>C</u>HD$_2$SOCD$_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMQC and gHMBCAD spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum. Coupling multiplicities are annotated as: s, singlet; d, doublet; dd, doublet of doublets; and t, triplet. ROESY correlations were observed using a mixing time of 250 ms. gHMBCAD correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



13

Position	**δ$_C$**	**Proton**	**δ$_H$ (*J* [Hz])**	**ROESY**	**gHMBC**
1-NH		1-NH	8.67 (s)	6-S-CH$_3$	3, 5
2	165.7				
3	65.0				
3-S-CH$_3$	13.0	3-S-CH$_3$	2.25 (s)		3
4-NH		4-NH	8.51 (s)	3-S-CH$_3$, 10-H	2, 6
5	165.3				
6	65.6				
6-S-CH$_3$	14.4	6-S-CH$_3$	2.14 (s)		6
7	64.6	7-H$_a$	3.45 (dd, $J_{7Ha,7Hb}$ = 11, $J_{7Ha,OH}$ = 6)	7-OH	5, 6
		7-H$_b$	3.65 (dd, $J_{7Ha,7Hb}$ = 11, $J_{7Hb,OH}$ = 6)	7-OH	
7-OH		7-OH	5.14 (t, *J* = 6)		
8	45.9	8-H$_a$	3.16 (d, $J_{8Ha,8Hb}$ = 15)		2, 3, 9, 14
		8-H$_b$	3.41 (d, $J_{8Ha,8Hb}$ = 15)	10-OH	10
9	122.0				
10	130.3	10-H	7.20 (d, *J* = 8)	8-H$_a$, 8-H$_b$	12
11	118.6	11-H	6.63 (t, *J* = 8)	10-H	9, 13
12	127.5	12-H	7.00 (t, *J* = 8)	11-H	10
13	114.9	14-H	6.76 (d, *J* = 8)	12-H, 10-OH,	11
14	155.3				
14-OH		14-OH	9.41 (s)	13-H	9, 14

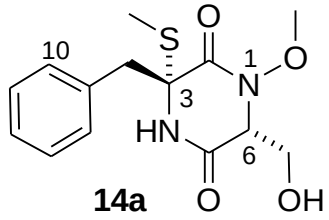
Table A.7: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **14a** in acetonitrile-d_3. Chemical shifts were referenced to δ(C<u>H</u>D$_2$CN) = 1.94 ppm and δ(^{13}<u>C</u>HD$_2$CN) = 1.3 ppm. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. ^{13}C chemical shifts were determined via HSQCAD and HMBC spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



14a

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-N				
1-N-O-CH_3	62.0	1-N-O-CH_3	3.52	
2	163.0			
3	68.8			
3-S-CH_3	13.4	3-S-CH_3	2.19	3
4-NH		4-NH	7.02	
5	165.2			
6	63.3	6-H	3.46 ($J_{6\text{-H},7\text{-Ha}}$ = 5, $J_{6\text{-H},7\text{-Hb}}$ = 5)	5
7	59.2	7-H_a	3.77 ($J_{7\text{-Ha},7\text{-OH}}$ = 6)	5
		7-H_b	3.78 ($J_{7\text{-Hb},7\text{-OH}}$ = 6)	
7-OH		7-OH	3.12	
8	46.9	8-H_a	3.00 ($J_{8\text{-Ha},8\text{-Hb}}$ = 13)	2, 3, 9, 10, 14
		8-H_b	3.51	2, 3, 9, 10, 14
9	135.3			
10, 14	131.7	10-H, 14-H	7.20	8, 12, 11, 10, 13, 14
11, 13	129.4	11-H, 13-H	7.30	9, 11, 13
12	128.5	12-H	7.28	

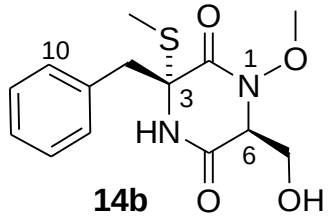
Table A.8: ^{1}H (600 MHz), and ^{15}N (62 MHz) spectroscopic data for compound **14a** DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm. (^{1}H,^{1}H)-Coupling constants were determined from the ^{1}H-NMR spectrum. Coupling multiplicity are annotated as: s, singlet; d, doublet; ddd, doublet of doublets of doublets; t, triplet; m, multiplet. ROESY correlations were observed using a mixing time of 250 ms. ^{15}N-gHMBCAD correlations (optimized for 8 Hz) are from the proton(s) stated to the indicated nitrogen.



14a

Position	**Proton**	**δ_H (*J* [Hz])**	**ROESY**	**^{15}N-gHMBC**	**δ ^{15}N**
1-N					185.4
1-N-O-CH_3		3.44 (s)	aromatic-H (7.18 ppm)	1-N	
2					
3					
3-S-CH_3		2.17 (s)	4-H, 8-H_a, 8-H_b, 7-OH		
4-NH	4-NH	9.00 (s)	8-H_a, aromatic-H (7.18 ppm)	4-NH (residual 1 bond)	130.0
5					
6		3.57 (t, *J* = 2)	7-H_a, 7-H_b, 7-OH		
7	7-H_a	3.62 (ddd, *J* = 11, *J* = 5, *J* = 3)	7-H_b		
	7-H_b	3.67 (ddd, *J* = 11, *J* = 5, *J* = 2)	7-H_a		
7-OH	7-OH	5.11 (t, *J* = 5)			
8	8-H_a	2.99 (d, *J* = 13)	8-H_b, aromatic-H (7.18 ppm, 7.22 ppm)		
	8-H_b	3.38 (d, *J* = 13)	8-H_a, aromatic-H (7.18 ppm, 7.22 ppm)	4-N	
9					
phenyl		7.28-7.19 (m)			

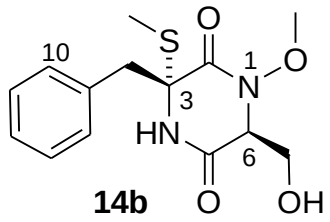
Table A.9: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **14b** in acetonitrile-d_3. Chemical shifts were referenced to δ($C\underline{H}D_2CN$) = 1.94 ppm and δ($^{13}\underline{C}HD_2CN$) = 1.3 ppm. ^{13}C chemical shifts were determined via HSQCAD and HMBC spectra. ($^{1}H,^{1}H$)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



14b

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-N				
1-N-O-CH_3	62.9	1-N-O-CH_3	3.74	
2	163.8			
3	68.2			
3-S-CH_3	13.1	3-S-CH_3	2.14	3
4-NH		4-NH	6.75	
5	165.9			
6	64.6	6-H	4.28 ($J_{6\text{-H},7\text{-Ha}}$ = 4, $J_{6\text{-H},7\text{-Hb}}$ = 3)	5
7	60.1	7-H_a	3.56 ($J_{7\text{-Ha},7\text{-Hb}}$ = 12, $J_{7\text{-Ha, 7-OH}}$ = 6)	
		7-H_b	3.61 ($J_{7\text{-Hb},7\text{-OH}}$ = 6)	
7-OH		7-OH	2.78	
8	44.9	8-H_a	3.11 ($J_{8\text{-Ha},8\text{-Hb}}$ = 14)	2, 3, 9, 10
		8-H_b	3.55	
9	135.3			
10,14	132.0	10-H, 14-H	7.24	8, 10, 14, 12
11,13	129.2	11-H, 13-H	7.30	
12	128.3		7.26	

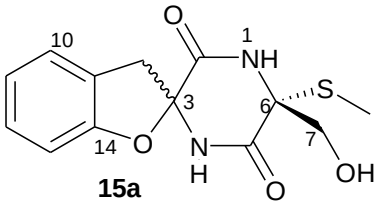
Table A.10: ^{1}H (600 MHz), and ^{15}N (62 MHz) spectroscopic data for compound **14b** DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm . (^{1}H,^{1}H)-Coupling constants were determined from the ^{1}H-NMR spectrum. Coupling multiplicity are annotated as: s, singlet; d, doublet; t, triplet; and m, multiplet. ROESY correlations were observed using a mixing time of 250 ms. ^{15}N-gHMBCAD correlations (optimized for 8 Hz) are from the proton(s) stated to the indicated nitrogen.



14b

Position	**Proton**	**δ_H (*J* [Hz])**	**ROESY**	**^{15}N-gHMBC**	**δ ^{15}N**
1-N					184.5
1-N-O-CH_3	1-N-O-CH_3	3.67 (s)	3-H, 6-H, 3-S-CH3,	1-N	
2					
3					
3-S-CH_3	3-S-CH_3	2.15 (s)	6-H		
4-NH		8.85 (s)	8-H_a, 8-H_b, aromatic-H (7.20 ppm)		
5					
6	6-H	4.32 (broad-t, *J* = 3)	7-H_a, 7-H_b		
7	7-H_a	3.51 (m)	7-H_b, 7-OH		
	7-H_b	3.51 (m)	7-H_a, 7-OH		
7-OH	7-OH	4.67 (t, *J* = 5) (No signal upon D2O addition)			
8	8-H_a	3.03 (d, *J* = 14)	aromatic-H (7.20 ppm)		
	8-H_b	3.49 (d, *J* = 14)	aromatic-H (7.20 ppm)		
9					
phenyl		7.23-7.19 (m)			

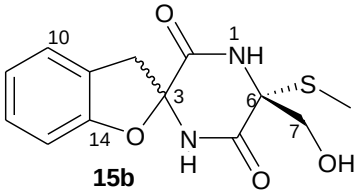
Table A.11: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **15a** (major diastereomer, diastereomeric ratio **15a**:**15b** = 6:4) in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via HMQC and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



15a

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-NH		1-NH	9.06	3, 5
2	164.7			
3	91.8			
4-NH		4-NH	9.61	2, 6
5	165.8			
6	66.2			
6-S-CH_3	12.2	6-S-CH_3	2.19	6
7	64.6	7-H_a	3.53 ($J_{7\text{-Ha},7\text{-Hb}}$ = 11, $J_{7\text{-Ha},7\text{-OH}}$ = 5)	5
		7-H_b	3.89 ($J_{7\text{-Hb},7\text{-OH}}$ = 6	
7-OH		7-OH	5.41	6
8	38.5	8-H_a	3.19 ($J_{8\text{-Ha},8\text{-Hb}}$ = 16)	2, 3, 9, 14
		8-H_b	3.83	2, 9, 14
9	125.4			
10	124.3	10-H	7.23 ($J_{10\text{-H},11\text{-H}}$ = 7)	12
11	120.9	11-H	6.89 ($J_{11\text{-H},12\text{-H}}$ = 7)	9, 13
12	127.7	12-H	7.13 ($J_{12\text{-H},13\text{-H}}$ = 7)	10, 14
13	108.7	13-H	6.77	9, 11
14	156.9	14-H		

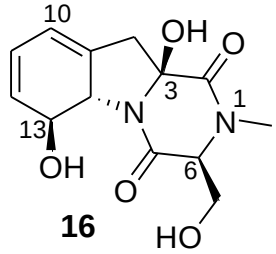
Table A.12: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **15b** (minor diastereomer, diastereomeric ratio **15a**:**15b** = 6:4) in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via HMQC and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



15b

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-NH		1-NH	8.99	3, 5
2	164.8			
3	91.5			
4-NH		4-NH	9.56	2, 6
5	165.1			
6	68.2			
6-S-CH_3	11.6	6-S-CH_3	2.07	6
7	64.6	7-H_a	3.52 ($J_{7\text{-Ha},7\text{-Hb}}$ = 11, $J_{7\text{-Ha},7\text{-OH}}$ = 5)	5
		7-H_b	3.89 ($J_{7\text{-Hb},7\text{-OH}}$ = 6	
7-OH		7-OH	5.54	
8	40.3	8-H_a	3.23 ($J_{8\text{-Ha},8\text{-Hb}}$ = 16)	2, 9, 14
		8-H_b	3.74	
9	125.5			
10	124.3	10-H	7.20 ($J_{10\text{-H},11\text{-H}}$ = 7)	12
11	120.8	11-H	6.88 ($J_{11\text{-H},12\text{-H}}$ = 7)	9, 13
12	127.9	12-H	7.13 ($J_{12\text{-H},13\text{-H}}$ = 7)	10, 14
13	108.5	13-H	6.76	9, 11
14	157.4	14-H		

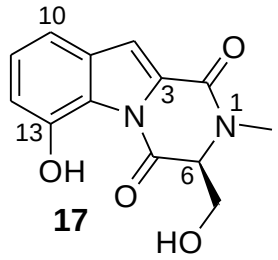
Table A.13: ^{1}H (900 MHz) and ^{13}C (227 MHz) NMR spectroscopic data for compound **16** in DMSO-d_6. Chemical shifts were referenced to $\delta(C\underline{H}D_2SOCD_3)$ = 2.50 ppm and $\delta(^{13}\underline{C}HD_2SOCD_3)$ = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBCAD spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum. Coupling multiplicities are annotated as: s, singlet; d, doublet; t, triplet; dt, doublet of triplets; and m, multiplet. gHMBCAD correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon. ROESY correlations were observed using a mixing time of 250 ms.



16

Position	**δ_C**	**Proton**	**δ_H (*J* [Hz])**	**gHMBC**	**ROESY**
1-N					
1-N-CH_3	32.0	1-N-CH_3	2.93	2, 6	6-H, 7-H
2	165.4				
3	88.9				
3-OH		3-OH	6.25 (d, *J* = 1)	8, 3	7-H_a, 7-H_b, 8-H_a
4-N					
5	168.6				
6	64.9	6-OH	4.21 (t, *J* = 3)	5	7-H_a, 7-H_b, 8-H_b
7	59.9	7-H_a	3.81 (dt, *J* = 4, *J* = 11)		7-OH
		7-H_b	3.91 (dt, *J* = 4, *J* = 11)		7-OH
7-OH		7-OH	6.47 (t, *J* = 4)	7, 6	
8	40.3	8-H_a	2.72 (d, *J* = 16)	3, 9, 10, 14	10-H, 14-H
		8-H_b	2.82 (d, *J* = 16)		10-H
9	117.8				
10	134.5	10-H	5.89 (m)		11-H
11	118.4	11-H	5.89 (m)		12-H
12	130.4	12-H	5.58 (m)	11	13-H
13	73.7	13-H	4.52 (d, *J* = 14)		13-OH
13-OH		13-OH	5.98 (s)	12, 13, 14	14-H
14	67.9	14-H	4.69 (d, *J* = 14)		

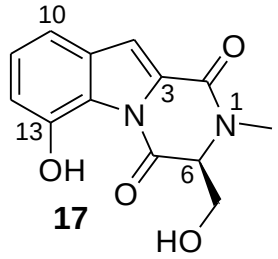
Table A.14: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **17** in acetonitrile-d_3. Chemical shifts were referenced to δ(C<u>H</u>D_2CN) = 1.94 ppm and δ(^{13}<u>C</u>HD_2CN) = 1.3 ppm. ^{13}C chemical shifts were determined via HSQCAD and HMBC spectra. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum or the dqfCOSY spectrum. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon.



17

Position	**δ_C**	**Proton**	**δ_H (J_{HH} [Hz])**	**HMBC**
1-N				
1-N-CH_3	31.6	1-N-CH_3	3.07	2, 6
2	157.8			
3	117.7 (DMSO-d_6)			
4-N				
5	166.0			
6	67.0	6-H	4.32 ($J_{6\text{-H},7\text{-Ha}}$ = 2, $J_{6\text{-H},7\text{-Hb}}$ = 3)	2, 5
7	62.4	7-H_a	3.99 ($J_{7\text{-Ha},7\text{-Hb}}$ = 12, $J_{7\text{-Ha},7\text{-OH}}$ = 6)	
		7-H_b	4.08 ($J_{7\text{-Hb},7\text{-OH}}$ = 6)	
7-OH		7-OH	3.35	
8	110.1	8-H	7.36	9, 10, 13, 14
9	119.1			
10	108.9	10-H	7.90 ($J_{10\text{-H},11\text{-H}}$ = 8)	9, 12
11	129.8	11-H	7.34 ($J_{11\text{-H},12\text{-H}}$ = 9)	13, 14
12	110.6	12-H	6.81	9, 10, 14
13	151.9			
13-OH			7.70	
14	137.0			

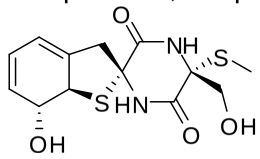
Table A.15: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **17** in DMSO-d_6. Chemical shifts were referenced to $\delta(C\underline{H}D_2SOCD_3)$ = 2.50 ppm and $\delta(^{13}\underline{C}HD_2SOCD_3)$ = 39.5 ppm. ^{13}C chemical shifts were determined via HMBC spectra and HSQCAD. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated carbon. (^{1}H,^{1}H)-coupling constants were determined from the ^{1}H-NMR spectrum. Coupling multiplicities are annotated as: s, singlet; d, doublet; t, triplet; dd, doublet of doublets; ddd, doublet of doublet of doublets; m, multiplet.



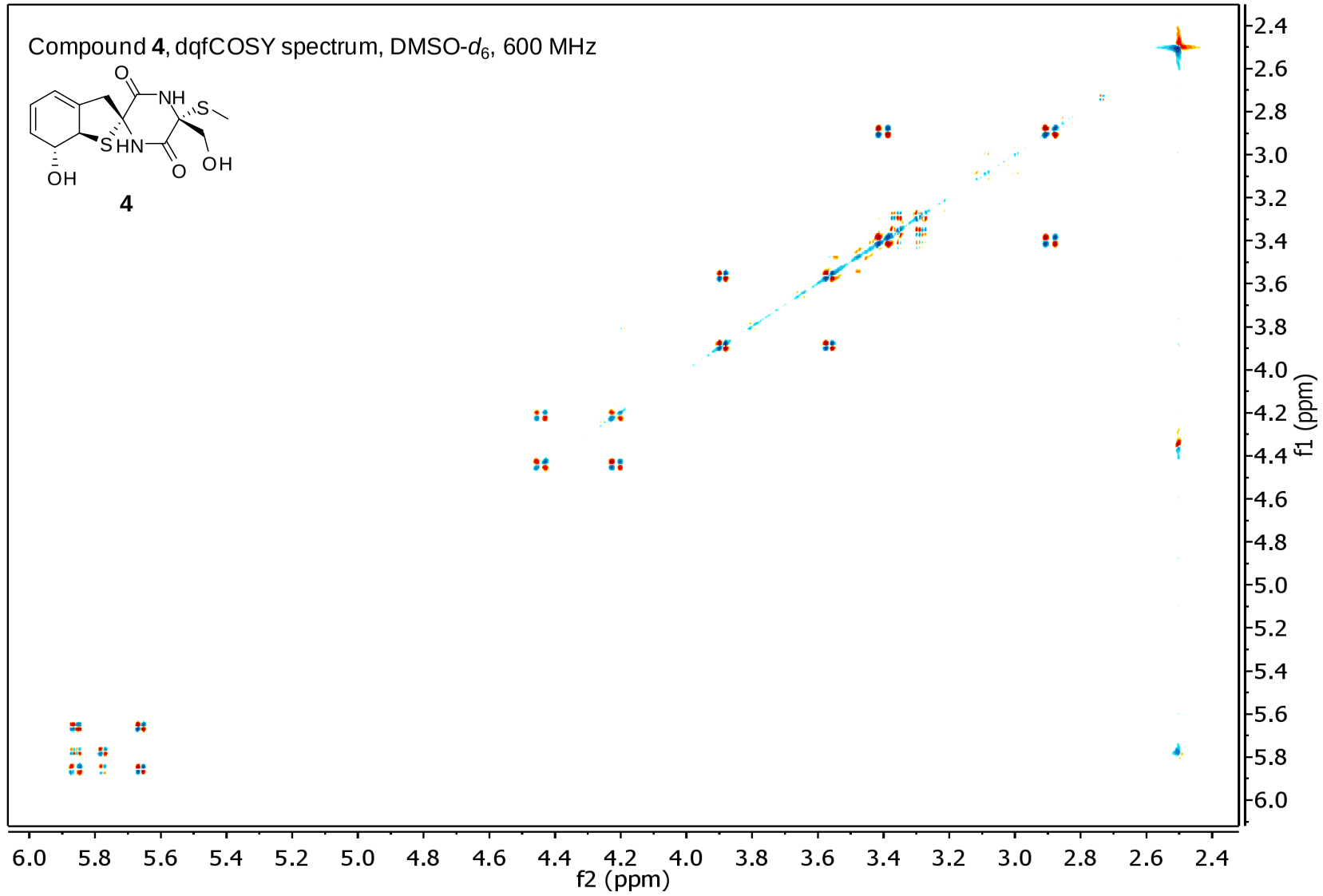
17

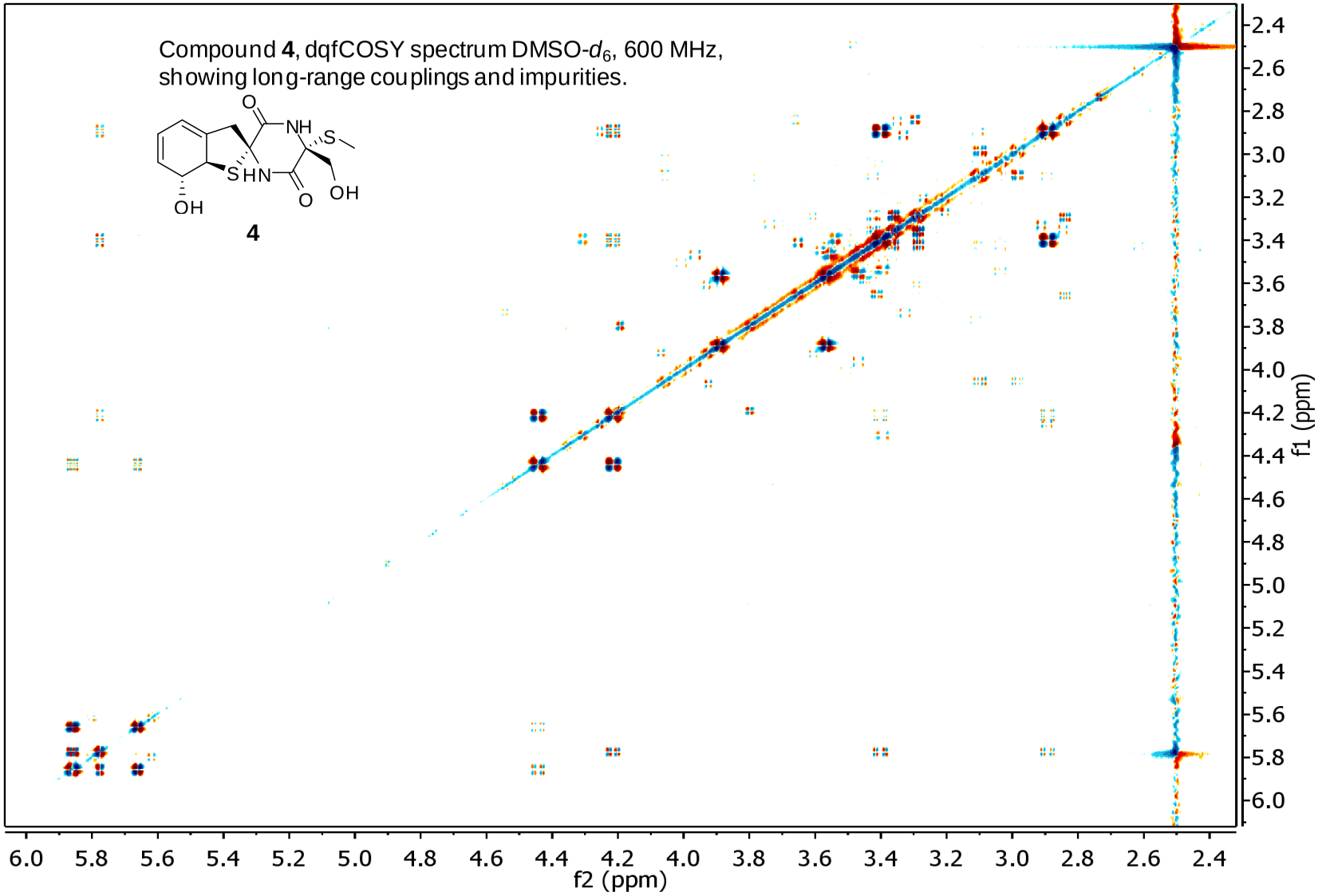
Position	**δ_C**	**Proton**	**δ_H (*J* [Hz])**	**HMBC**
1-N				
1-N-CH_3	30.5	1-N-CH_3	3.02 (s)	2, 6
2	156.3			
3	117.7			
4-N				
5	165.1			
6	65.8	6-H	4.49 (broad-t, J = 2)	5
7a	60.4	7-H_a	3.85 (ddd, J = 12, J = 6, J = 2)	
		7-H_b	3.98 (ddd, J = 11, J = 5, J = 3)	
7-OH		7-OH	5.36 (t, J = 6)	
8	108.7	8-H	7.33 (d, J < 1)	3, 11, 14
9	118.1			
10	106.6	10-H	7.79 (d, J = 8)	9, 12
11	128.5	11-H	7.31 (t, J = 8)	13, 14
12	109.7	12-H	6.76 (dd, J = 8, J < 1)	9, 10, 14
13	152.0			
13-OH			10, 19 (s)	9, 12, 13, 14
14	135.4			

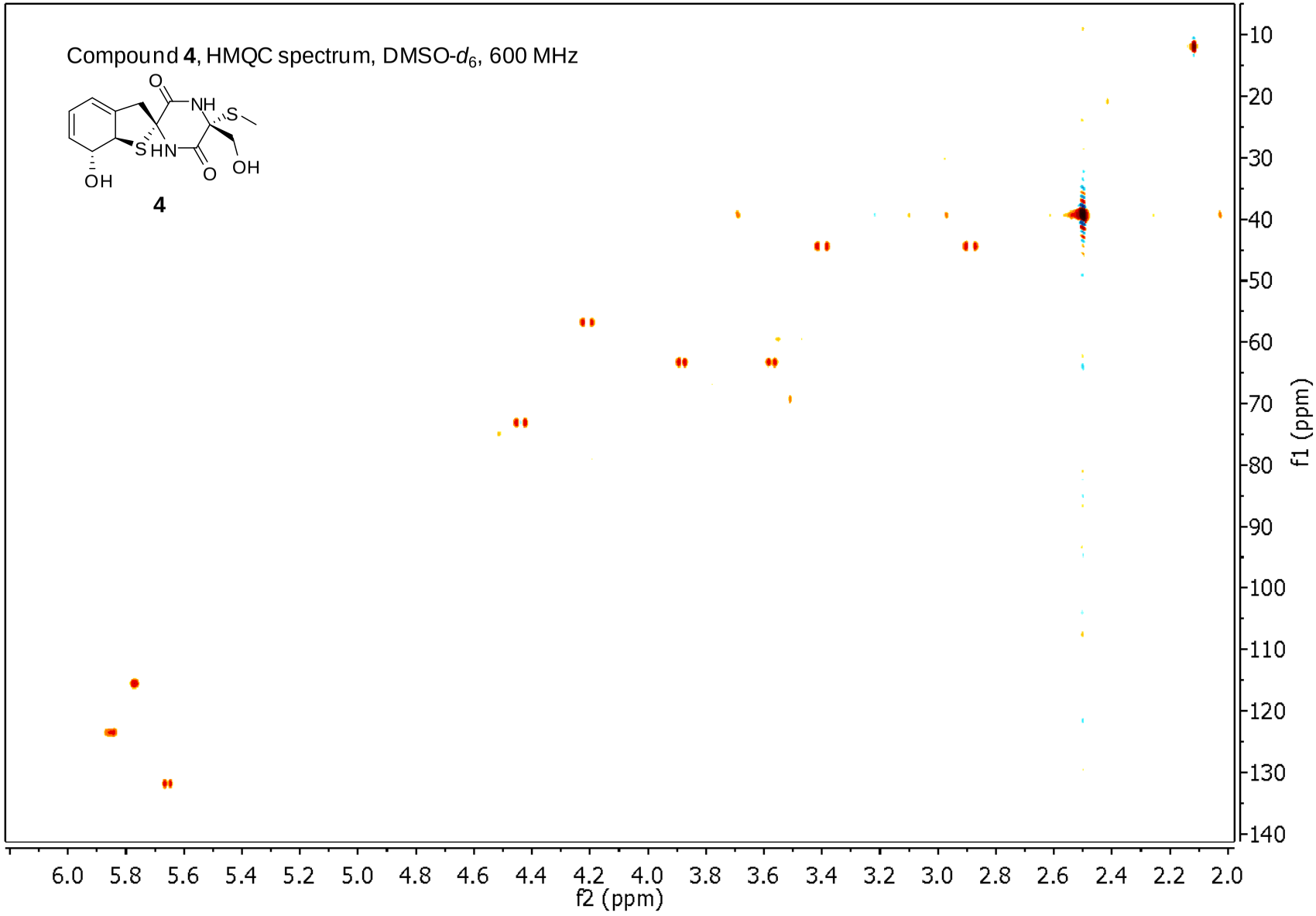
Compound **4**, ^{1}H spectrurm, DMSO-d_6, 600MHz

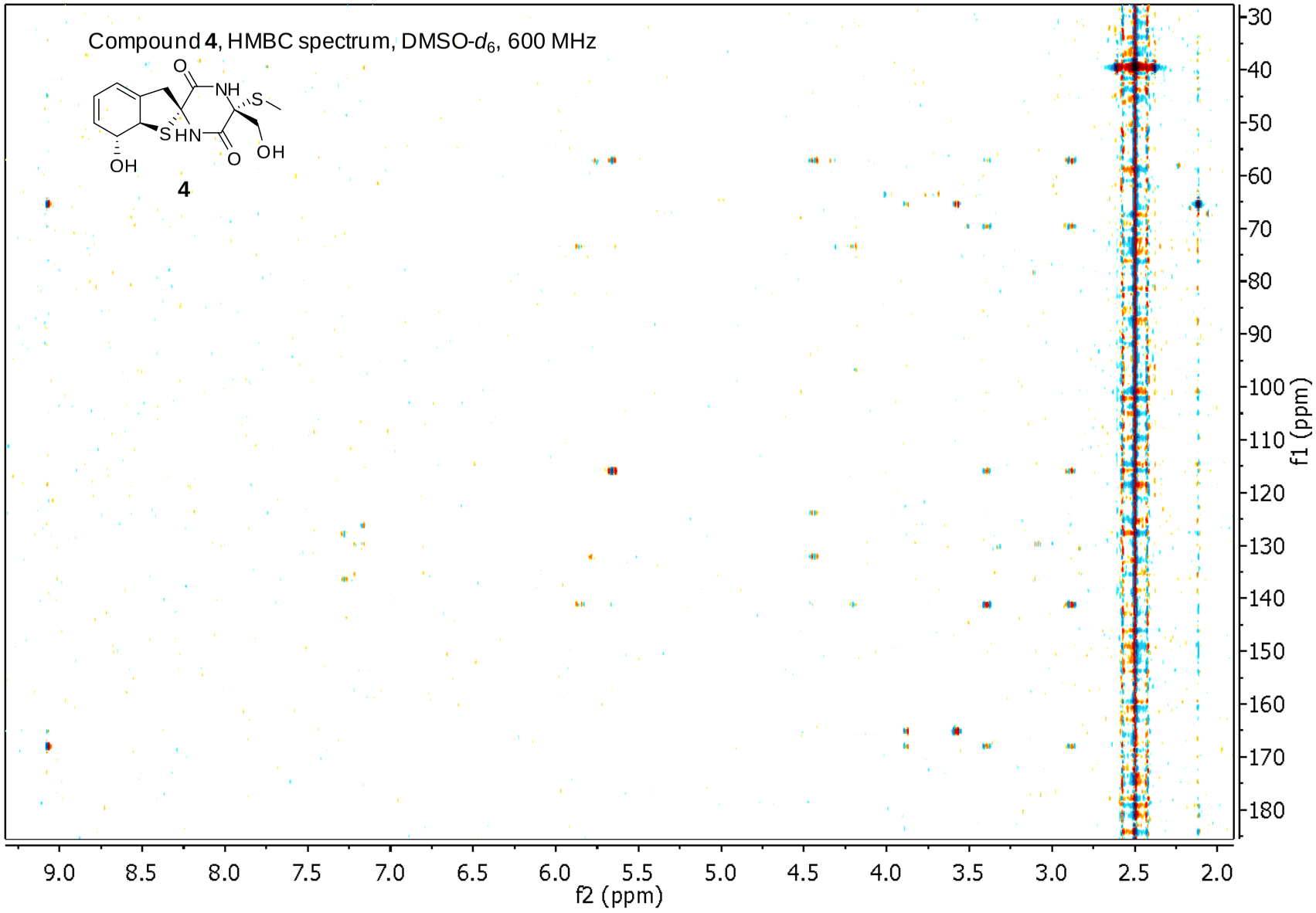


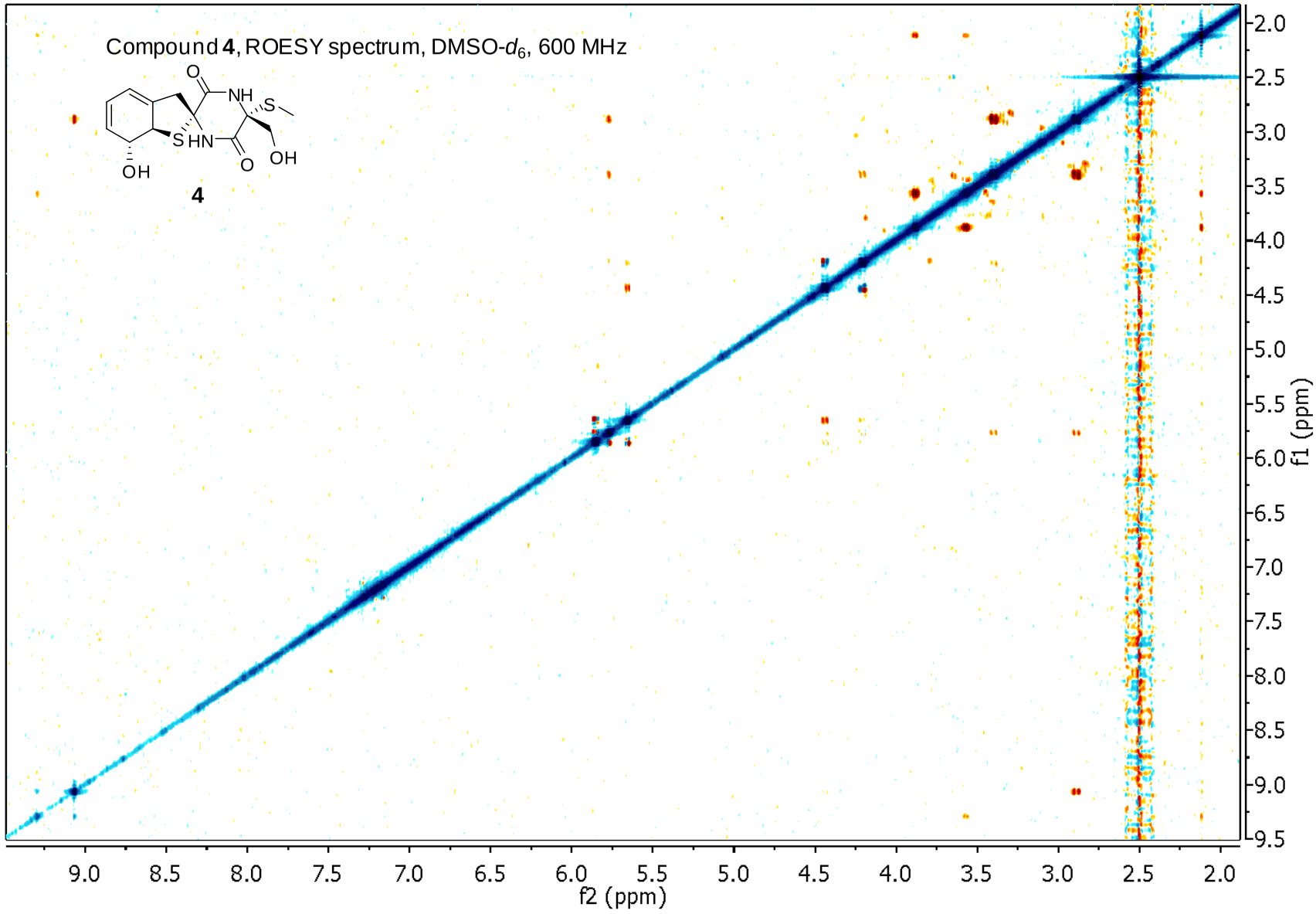
4

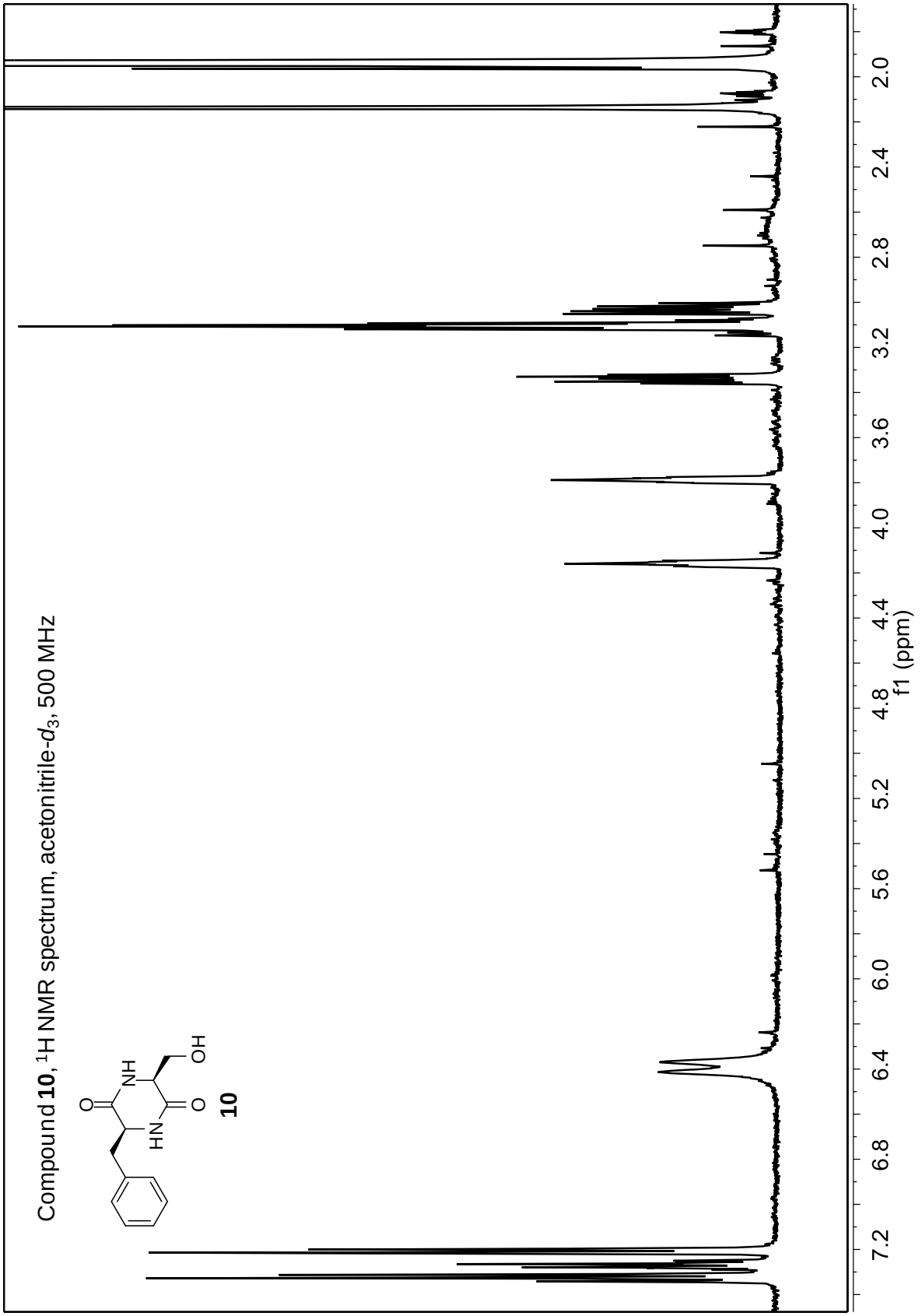

Compound 4, dqfCOSY spectrum, DMSO-d_6, 600 MHz
O
NH
S
S
HN
O
OH
OH
4
f2 (ppm)
f1 (ppm)
6.0 5.8 5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8 3.6 3.4 3.2 3.0 2.8 2.6 2.4


Compound 4, dqfCOSY spectrum DMSO-d_6, 600 MHz,
showing long-range couplings and impurities.
O
NH
S
S
HN
O
OH
OH
4
6.0 5.8 5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8 3.6 3.4 3.2 3.0 2.8 2.6 2.4
f2 (ppm)
2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0 4.2 4.4 4.6 4.8 5.0 5.2 5.4 5.6 5.8 6.0
f1 (ppm)

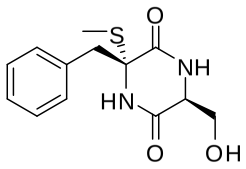

Compound 4, HMQC spectrum, DMSO-d_6, 600 MHz
O
NH
S
S
HN
O
OH
OH
4
f2 (ppm)
f1 (ppm)
6.0 5.8 5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8 3.6 3.4 3.2 3.0 2.8 2.6 2.4 2.2 2.0
10 20 30 40 50 60 70 80 90 100 110 120 130 140


Compound 4, HMBC spectrum, DMSO-d_6, 600 MHz
O
NH
S
S
HN
O
OH
OH
4
f2 (ppm)
f1 (ppm)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180


Compound 4, ROESY spectrum, DMSO-d_6, 600 MHz
O
NH
S
S
HN
O
OH
OH
4
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
f2 (ppm)
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
f1 (ppm)

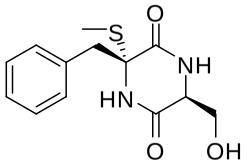

Compound **10**, ^{1}H NMR spectrum, acetonitrile-d_3, 500 MHz
10
7.2
6.8
6.4
6.0
5.6
5.2
4.8
4.4
4.0
3.6
3.2
2.8
2.4
2.0
f1 (ppm)

Compound **12**, ^{1}H NMR spectrum, acetonitrile-d_3, 600 MHz

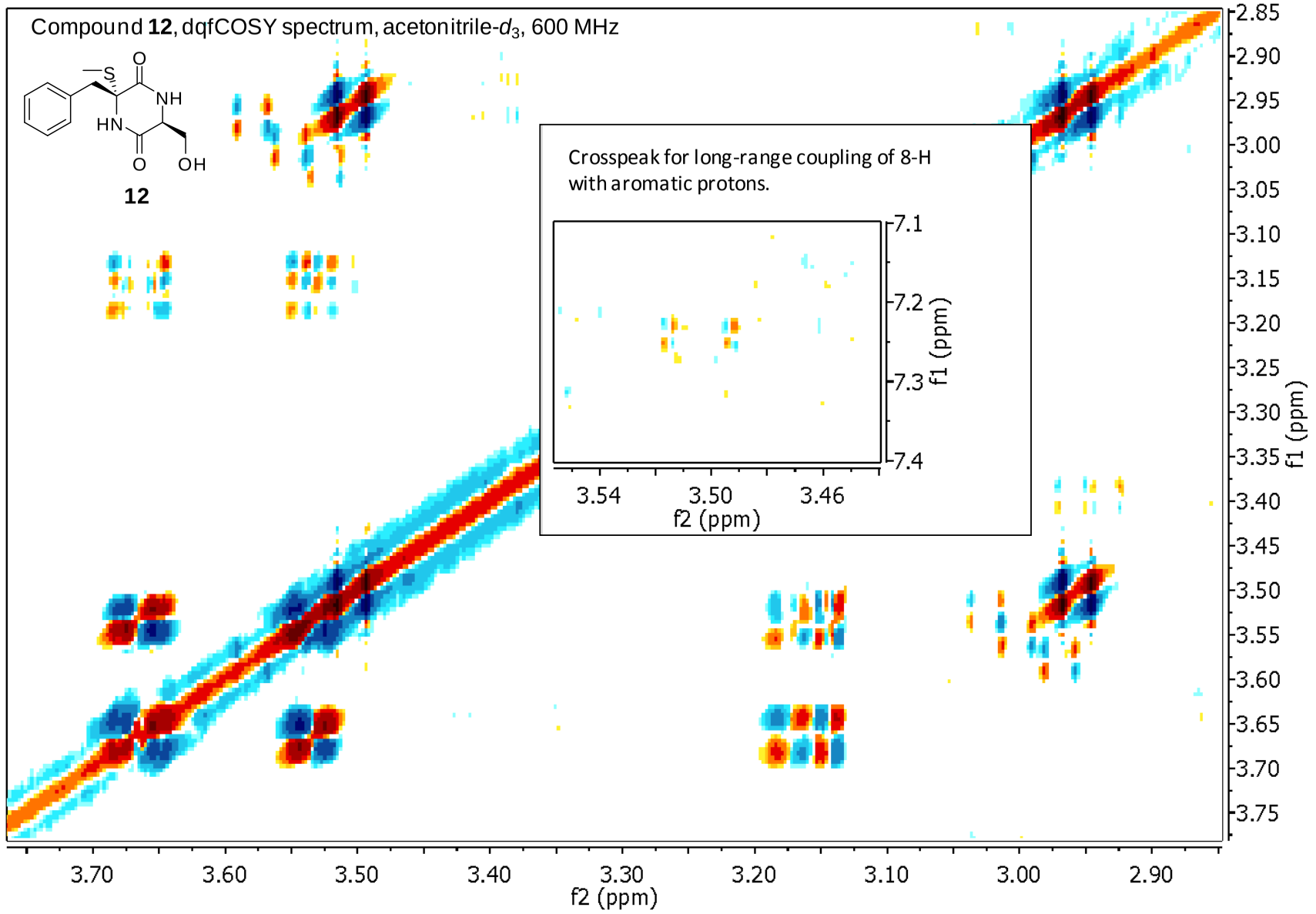


12

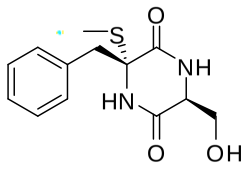
Compound **12**, ^{1}H NMR spectrum, acetonitrile-d_3, 600 MHz



12

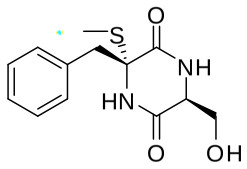

Compound 12, dqfCOSY spectrum, acetonitrile-d_3, 600 MHz
12
Crosspeak for long-range coupling of 8-H with aromatic protons.
f2 (ppm)
f1 (ppm)
3.54
3.50
3.46
7.1
7.2
7.3
7.4
3.70
3.60
3.50
3.40
3.30
3.20
3.10
3.00
2.90
2.85
2.90
2.95
3.00
3.05
3.10
3.15
3.20
3.25
3.30
3.35
3.40
3.45
3.50
3.55
3.60
3.65
3.70
3.75

Compound **12**, HMBC spectrum, acetonitrile-d_3, 600 MHz

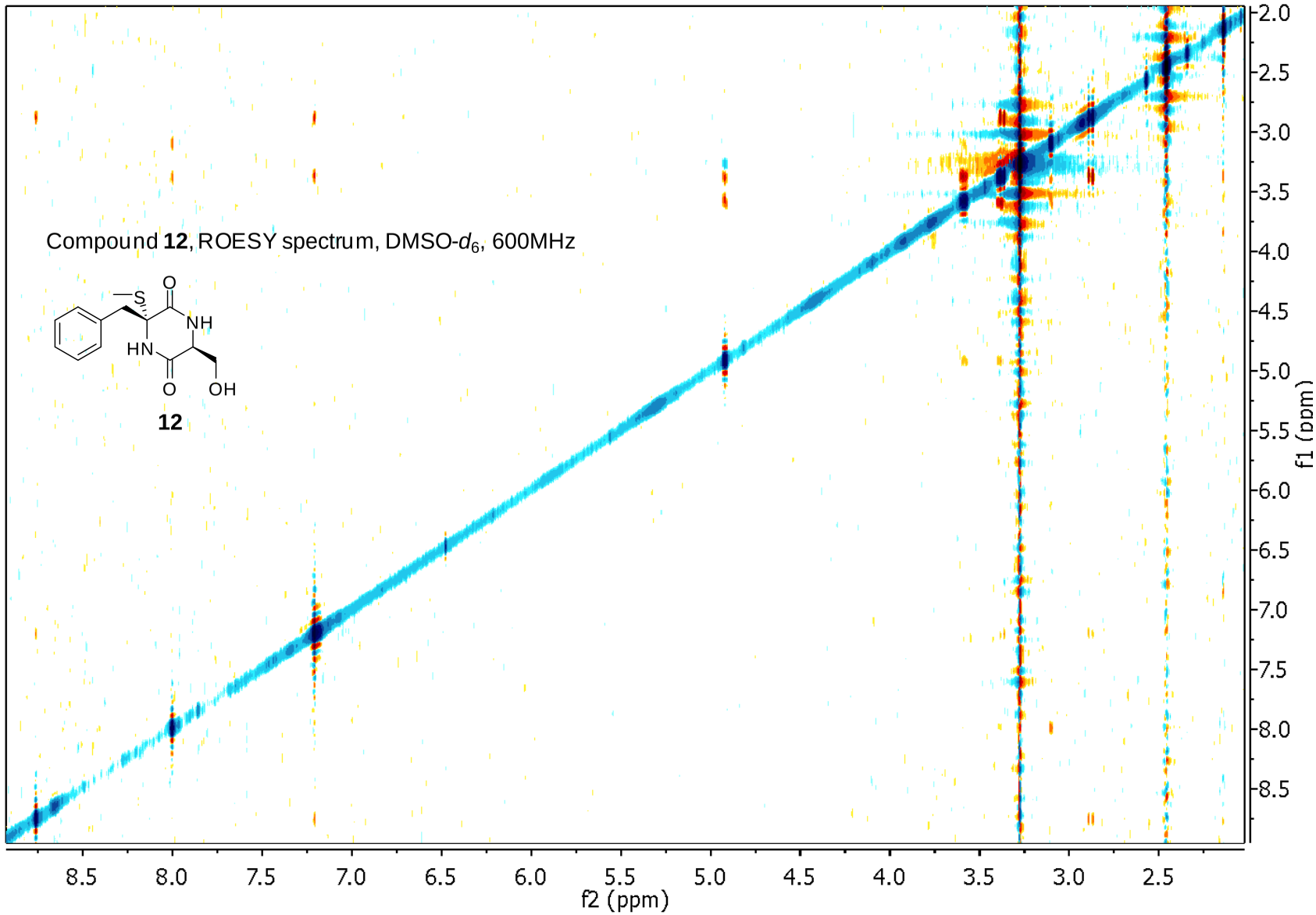


12

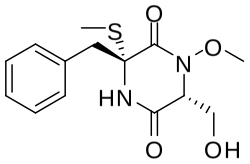
Compound **12**, HMBC spectrum, acetonitrile-d_3, 600 MHz



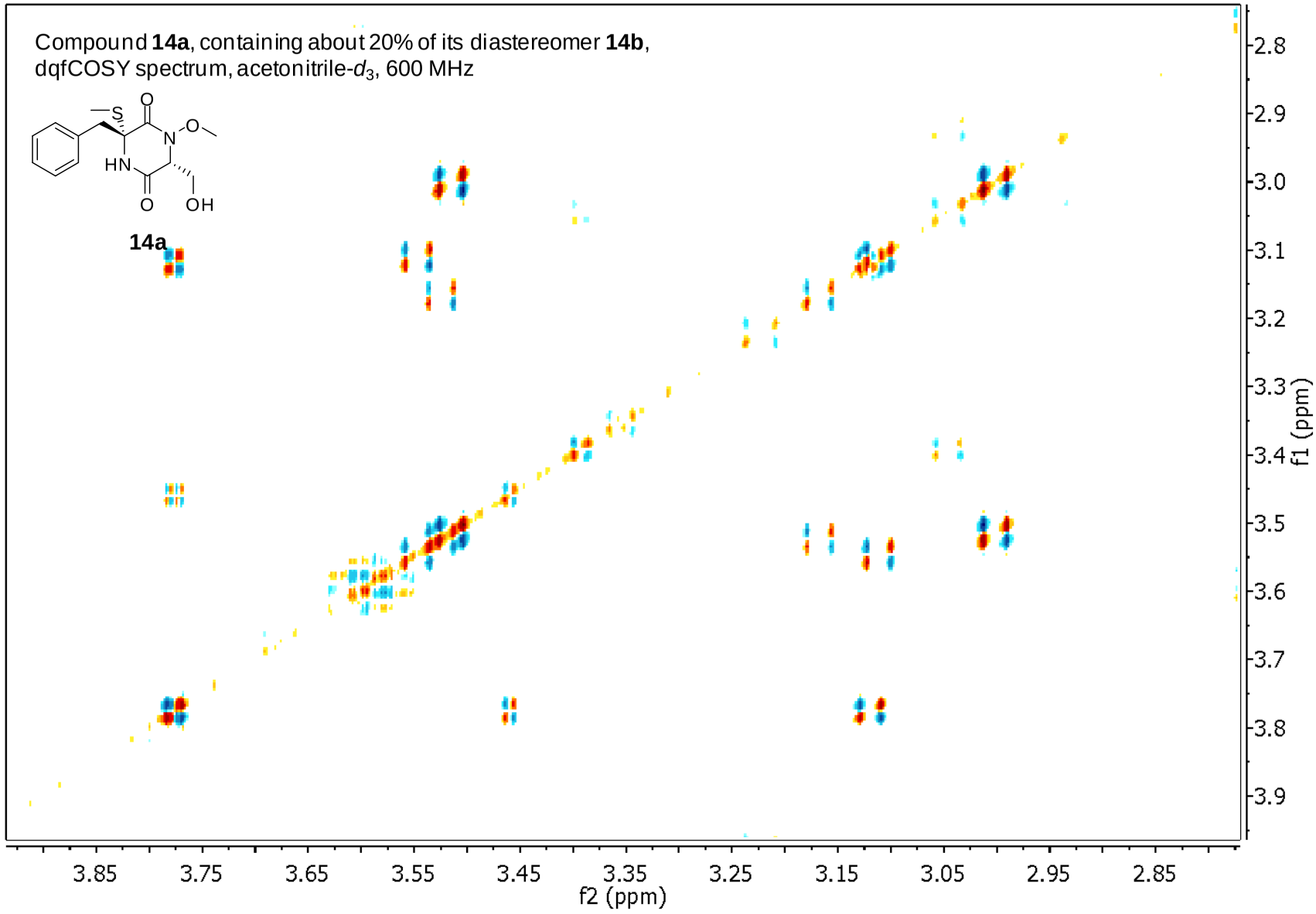
12


Compound 12, ROESY spectrum, DMSO-d_6, 600MHz
S
O
NH
HN
O
OH
12
f2 (ppm)
f1 (ppm)

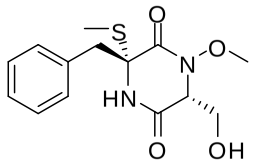
Compound **14a**, ^{1}H NMR spectrum, acetonitrile-d_3, 600 MHz



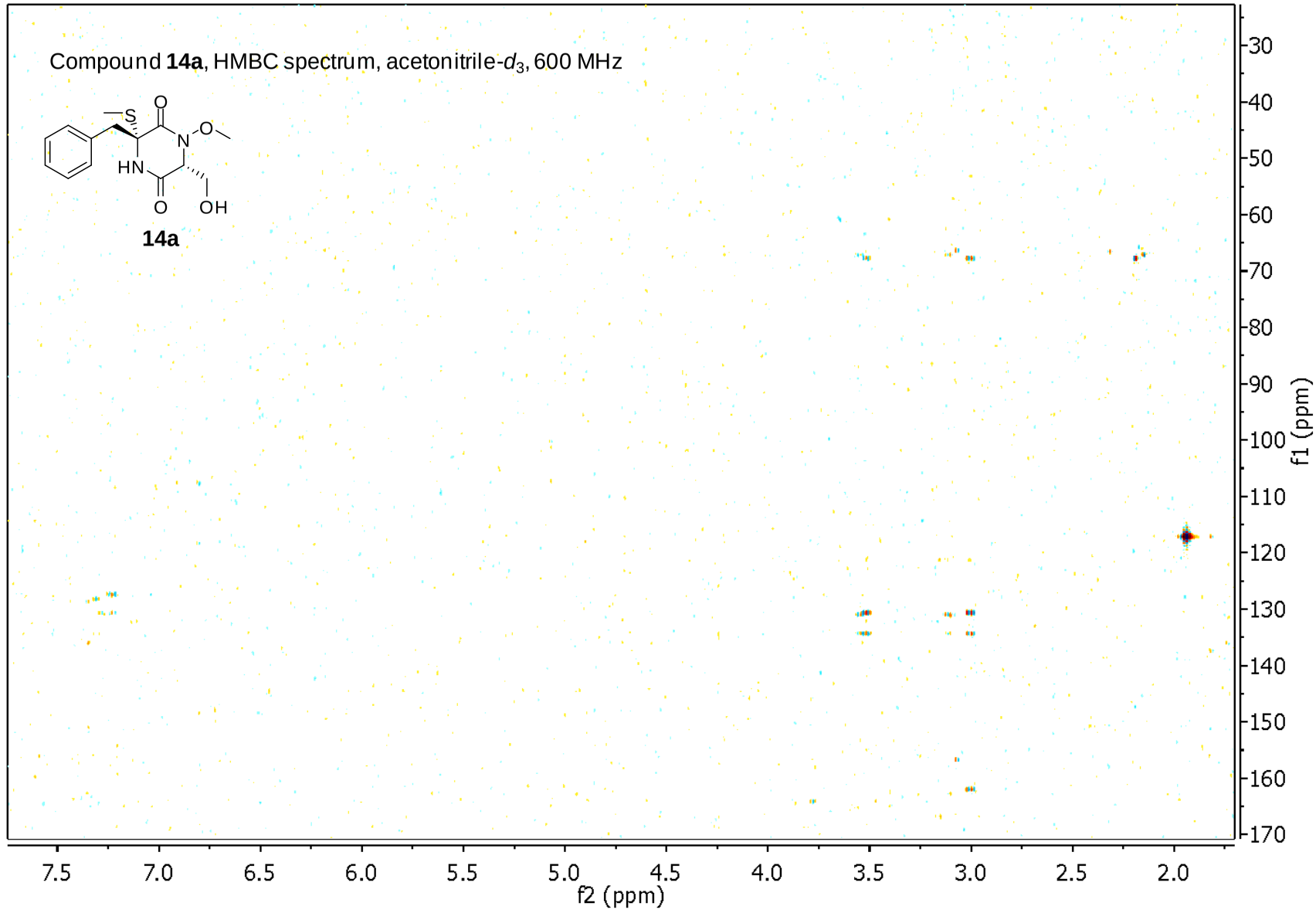
14a

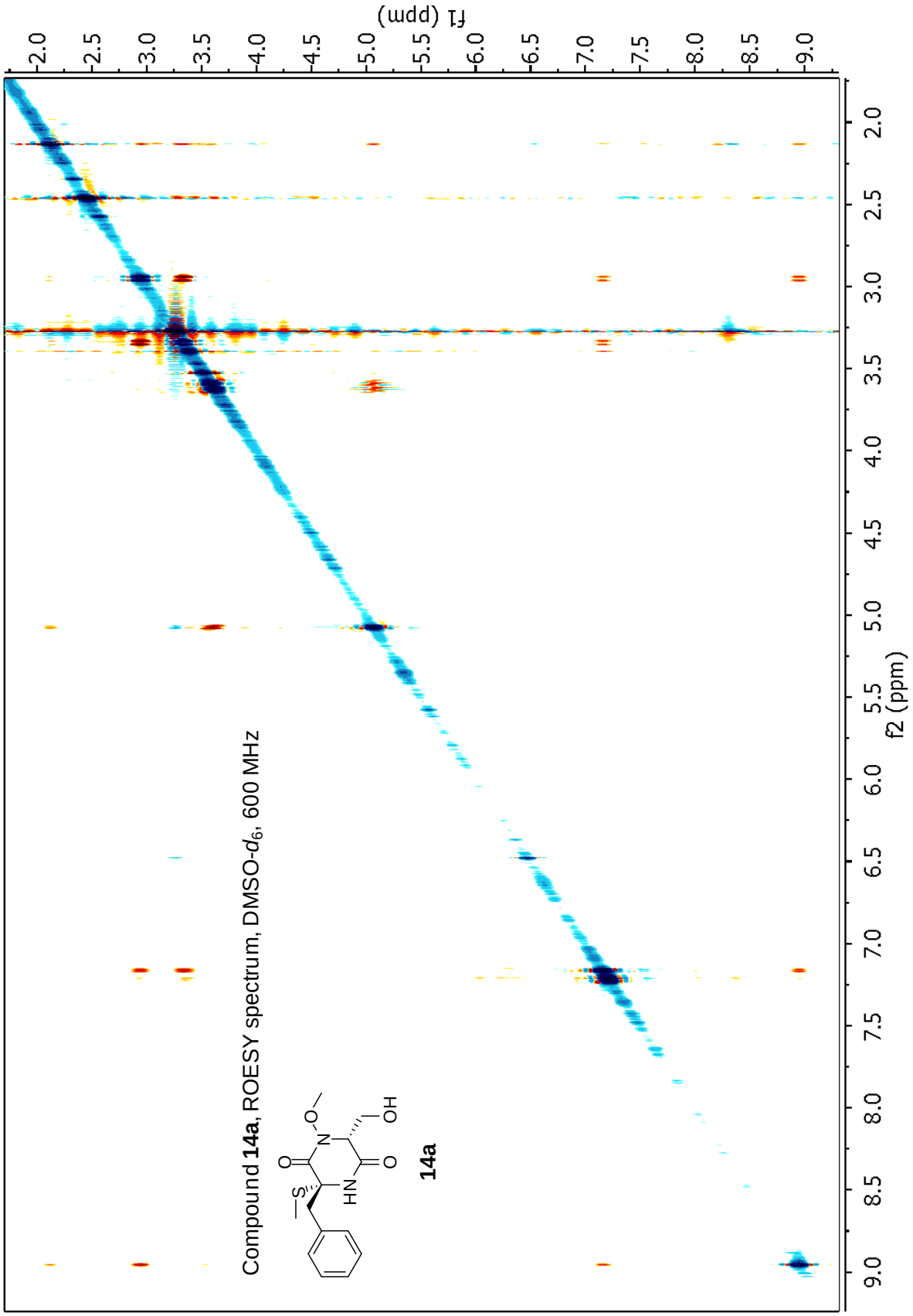

Compound **14a**, containing about 20% of its diastereomer **14b**,
dqfCOSY spectrum, acetonitrile-d_3, 600 MHz
S
O
N
O
HN
O
OH
14a
2.8
2.9
3.0
3.1
3.2
3.3
3.4
3.5
3.6
3.7
3.8
3.9
f1 (ppm)
3.85
3.75
3.65
3.55
3.45
3.35
3.25
3.15
3.05
2.95
2.85
f2 (ppm)

Compound **14a**, HSQC spectrum, acetonitrile-d_3, 600 MHz

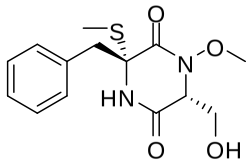


14a


Compound **14a**, HMBC spectrum, acetonitrile-d_3, 600 MHz
14a
f2 (ppm)
f1 (ppm)


Compound 14a, ROESY spectrum, DMSO-d6, 600 MHz
14a
f1 (ppm)
f2 (ppm)

Compound **14a**, (^{1}H, ^{15}N)-HMBC spectrum, DMSO-d_6, 600 MHz



14a

REFERENCES

(1) Shimizu, K.; Keller, N. P. *Genetics* **2001**, *157*, 591.

(2) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *PNAS USA* **2009**, *106*, 7708.

(3) Schroeder, F. C.; Gibson, D. M.; Churchill, A. C.; Sojikul, P.; Wursthorn, E. J.; Krasnoff, S. B.; Clardy, J. *Angew Chem Int Ed Engl* **2007**, *46*, 901.

(4) Balibar, C. J.; Walsh, C. T. In *Biochemistry*; **2006**, *45*, 5029.

Appendix B

A BACKUP PLAN FOR SELF-PROTECTION: S-METHYLATION OF HOLOMYCIN BIOSYNTHETIC INTERMEDIATES IN STREPTOMYCES CLAVULIGERUS

1. General Analytical Methods and Equipment: **(A) NMR-spectroscopy.** NMR spectroscopic instrumentation: a Varian INOVA 600 MHz NMR spectrometer (600 MHz reference frequency for ^{1}H, 151 MHz for ^{13}C) equipped with an HCN indirect-detection probe. ^{1}H,^{13}C-(g)HMBC spectra were acquired using the following parameters: 0.25 s acquisition time, 256 increments and 64-128 scans per increment. PW_{90} was determined for each sample individually. Spectra were optimized for $J_{H,C}$ = 6 Hz. Gradient and non-gradient ^{1}H,^{13}C-HMQC[AD] spectra were acquired using the following parameters: 0.25 s acquisition time, 300-600 increments, 32-64 scans per increment. ROESY spectra were acquired using the following parameters: 0.25 s acquisition time, 0.275 s mixing time, 200 increments, 64 scans per increment. Susceptibility-matched NMR tubes (Shigemi) were used for sample amounts smaller than 5 mg. NMR spectra were processed and baseline corrected using Varian VNMR and MestReC and MestReNova software packages.

(B) Mass spectrometry. High-resolution mass spectrometry (HRMS) was performed on a Waters nanoACQUITY UPLC system equipped with a Waters Acquity UPLC HSS C18 column (2.1 x 100 mm, 1.8 µm particle diameter) connected to a Xevo G2 QTof Mass Spectrometer. For HPLC-MS analysis, the eluent of an Agilent 1100 HPLC system, equipped with a diode array detector, was injected into a Quattro II spectrometer (Micromass/Waters) operated in positive-electrospray ionization mode (ESI^{+}). HPLC/ESI^{+}-MS data acquisition and processing was controlled by Waters MassLynx software.

(C) Chromatography. Flash chromatography was performed using a Teledyne ISCO CombiFlash system. An Agilent Zorbax Eclipse XDB-C8 column (4.6 x 150 mm, 5

μm particle diameter) was used in the HPLC/ESI^+-MS mutant profiling analysis and the crude Δ*hlmI*/ΔORF15 TCEP reduction experiment.

2. Bacterial Strains and Experimental Growth Conditions: Bacterial strains used in this study include *S. clavuligerus* wild-type, ΔORF15, and Δ*hlmI*/ΔORF15. The wild-type and ΔORF15 were provided by Dr. Kapil Tahlan (Memorial University of Newfoundland, Canada). The mutant strain Δ*hlmI*/ΔORF15 was generated in previous work[1]. *S. clavuligerus* wild-type and mutants were grown in parallel under the same conditions. The seed cultures were started using frozen spore stocks of each strain and grown in 250 mL flasks containing 50 mL Tryptic Soy Broth (TSB) at 30 °C for 24-36 h until dense. The $OD_{600\ nm}$ was measured after a 16-fold dilution. The amount of seed culture for inoculation was calculated to achieve an inoculating $OD_{600\ nm}$ of 0.25 in a 50 mL volume. The normalized amount of seed cultures was spun down and the pellets were washed and used to inoculate 50 mL of production medium modified GSPG in 20 × 250 mL flasks for a 500 mL-scale growth. The modified GSPG medium contains the following ingredients in one liter: glycerol, 5 g; proline, 2.5 g; glutamic acid, 1.5 g; NaCl, 5 g; K_2HPO_4, 2 g; $CaCl_2$, 0.4 g; $MnCl_2•4H_2O$, 0.1 g; $FeCl_3•6H_2O$, 0.1 g; $ZnCl_2$, 0.05 g; $MgSO_4•7H_2O$, 1 g; and distilled water at pH 7. The production cultures were incubated at 30 °C for 48 h with shaking. The resulting cultures were harvested as a mixture of cells and culture supernatant and frozen at −20 °C.

3. Culture Extraction, Sample Enrichment, and NMR Sample Preparation: Whole liquid cultures of the various *S. clavuligerus* mutants examined were frozen and lyophilized. The residue was extracted with a 10% water in acetonitrile solution v/v (1 mL extraction solution per 5 mL liquid culture). The resulting extract was filtered over acetonitrile- and water-washed cotton. The extraction process was repeated twice more (1 mL extraction solution per 10 mL liquid culture). The extracts were evaporated to

dryness and absorbed onto 3 g of octadecyl-functionalized silica gel. The samples were then fractionated on a 100 g RediSep Reverse-Phase (RP) C18 Gold flash chromatography column using a water-acetonitrile solvent gradient (flow rate 75 mL/min), starting with 5 min of 5% acetonitrile followed by a linear increase to 20% acetonitrile over 12.5 min, followed by a linear increase to 65% acetonitrile over 5 min, followed by a linear increase to 100% acetonitrile over 2 min, and ending with 6 min of 100% acetonitrile isocratic elution. Fractions were combined into three pools; pool 1, 5-12% acetonitrile; pool 2, 12-32% acetonitrile; pool 3, 32-100% acetonitrile. Pools were evaporated to dryness using rotary evaporation *in vacuo* at room temperature. Pool 1 samples were dissolved in 0.6 mL 99.9% d DMSO-d_6 and placed in an NMR tube. Pool 2 and pool 3 samples were suspended in ~0.2 mL of acetonitrile-d_3. The resulting suspensions were centrifuged to remove insoluble materials; this was repeated two additional times, and the supernatant was subjected to NMR spectroscopic analysis.

4. Pool 1 Fractionation: To enrich a sample of compound **3**, the pool 1 Δ*hlmI*/ΔORF15 sample (see previous section, Appendix B) was subjected to flash chromatography. The sample was dissolved in a 10% water/acetonitrile (v/v) solution and silica gel was added to the solution. Solvent was removed *in vacuo* using rotary evaporation. The silica gel-adsorbed sample was transferred to a loading column. Using a CombiFlash system equipped with a 24 g RediSep column, the sample was chromatographed utilizing a 10% water in acetonitrile/dichloromethane (v/v) gradient (flow rate 35 mL/min) starting with 0% water-acetonitrile solution for 2 min, followed by a linear increase to 10% water-acetonitrile solution over 3 min, followed by a linear increase to 20% water-acetonitrile solution over 2 min, followed by a linear increase to 70% water-acetonitrile solution over 4 min, followed by a linear increase to 100% water-acetonitrile solution over 1 min, and ending with 6 min of 100% water-acetonitrile. The ^{1}H spectrum for fraction 58 (eluting at 14 min, 100% water-acetonitrile) indicated the presence of **3**.

5. Pool 2 Fractionation: To enrich samples of **10** and **11** the pool 2 Δ*hlmI*/ΔORF15 sample (see above) was subjected to flash chromatography. The sample was dissolved in water and deposited onto a loading column containing 4 g of octadecyl-functionalized silica gel. Using a CombiFlash system equipped with a 13 g RediSep C18 column, the sample was chromatographed utilizing a water (containing 0.1% acetic acid)/acetonitrile gradient (flow rate 30 mL/min) starting with 5% acetonitrile for 0.9 min, followed by a linear increase to 20% acetonitrile over 3 min, followed by a linear increase to 55% acetonitrile over 0.9 min, followed by a linear increase to 100% acetonitrile over 0.4 min, and ending with 0.9 min of 100% acetonitrile. The ^{1}H spectra acquired for pooled fractions 12-15 (17-19% acetonitrile) and fractions 18-23 (32-65% acetonitrile) showed the presence of compounds **10** and **11**, respectively.

6. HPLC-ESI^{+}-MS Mutant Extract Profiling and Δhlml/ΔORF15 TCEP-Reduction Experiment: Unless otherwise specified, samples utilized in our HPLC/ESI^{+}-MS studies were prepared by suspending the crude bacterial extract in 1 mL of a 9:1 mixture of water and acetonitrile (v/v), followed by filtration of the extract through cotton. 3 µL of sample was injected into an Agilent 1100 series HPLC equipped with an Agilent Zorbax Eclipse XDB-C8 column. A water (0.1% acetic acid)/acetonitrile gradient starting with 5% acetonitrile (v/v) for 5 min, followed by a linear increase to 100% acetonitrile over 15 min and ending with 10 min of 100% acetonitrile, was used as mobile phase. The HPLC eluent was split (1:10) into a Quattro II triple quadrupole mass spectrometer operated in ESI^{+} mode. HPLC/ESI^{+}-MS data were processed in MassLynx software, and peak areas were measured using a peak-to-peak amplitude of 20 with no smoothing applied (see Appendix B **Figure B.4** for ion chromatograms corresponding to compounds **3**, **5**, **6**, and **8**-**11**.)

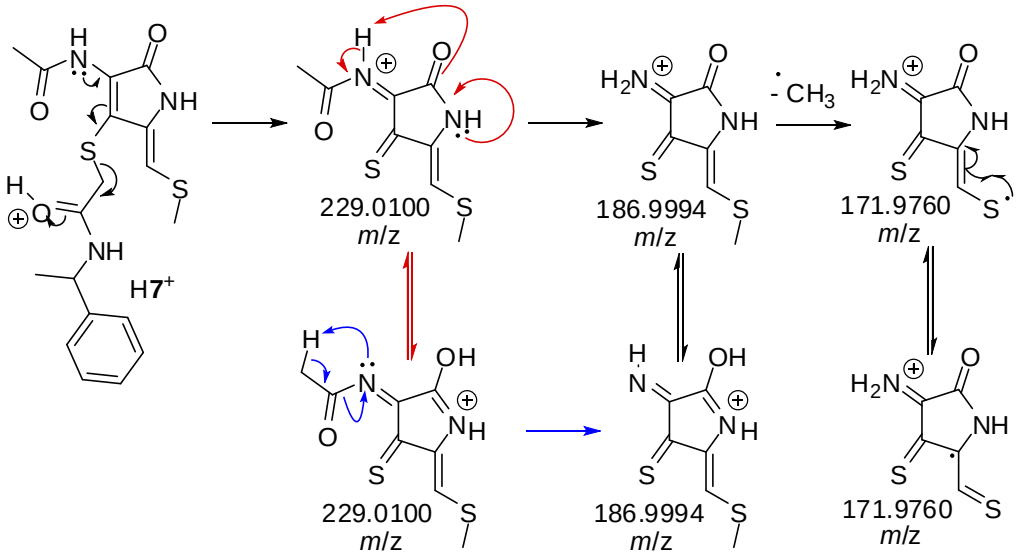
7. MS and MSMS Analysis of Compounds 3, and 5-11: **(A) HRMS, compounds 3, and 5-11.** Enriched samples of **3**, and **5**-**11** were dissolved in 200–500 µL of either 10%

water in acetonitrile (v/v) or 100% acetonitrile and injected into the UPLC/ESI$^+$-HRMS system described above. Various chromatographic solvent gradients were used; all using water (0.1% formic acid) and acetonitrile (0.1% formic acid) as solvents. Mass spectra were acquired in ESI$^+$ mode.

Compound	**Observed *m/z***	**Ion**	**Ion Formula**	**Calculated *m/z***
3	231.0257	$[M + H]^+$	$C_8H_{11}N_2O_2S_2^+$	231.0256
3	253.0081	$[M + Na]^+$	$C_8H_{10}N_2NaO_2S_2^+$	253.0076
3	189.0160	$[M + H - C_2H_2O]^+$	$C_6H_9N_2OS_2^+$	189.0151
5	203.0309	$[M + H]^+$	$C_7H_{11}N_2OS_2$	203.0307
6	385.0456	$[M + H]^+$	$C_{14}H_{17}N_4O_3S_3^+$	385.0457
7	392.1107	$[M + H]^+$	$C_{18}H_{22}N_3O_3S_2^+$	392.1097
8	459.0284	$[M + H]^+$	$C_{16}H_{19}N_4O_4S_4^+$	459.0284
8	481.0095	$[M + Na]^+$	$C_{16}H_{18}N_4NaO_4S_4^+$	481.0103
8	229.0103	$[M - {}^{\cdot}C_8H_9N_2O_2S_2]^+$, heterolytic S-S bond cleavage,	$C_8H_9N_2O_2S_2^+$	229.0105
8	231.0260	$[M + 2H - C_8H_9N_2O_2S_2]^+$, homolytic S-S bond cleavage	$C_8H_{11}N_2O_2S_2^+$	231.0256
9	438.9987	[M + Na]+	$C_{14}H_{16}N_4NaO_3S_4^+$	438.9997
9	231.0248	$[M + 2H - C_6H_7N_2OS_2]^+$, homolytic S-S bond cleavage	$C_8H_{11}N_2O_2S_2^+$	231.0256
10	245.0418	$[M + H]^+$	$C_9H_{13}N_2O_2S_2^+$	245.0413
10	267.0233	$[M + Na]^+$	$C_9H_{12}N_2NaO_2S_2^+$	267.0232
10	203.0306	$[M + H - C_2H_2O]^+$	$C_7H_{11}N_2OS_2^+$	203.0307
11	219.0259	$[M + H]^+$	$C_7H_{11}N_2O_2S_2^+$	219.0256
11	241.0079	$[M + Na]^+$	$C_7H_{10}N_2NaO_2S_2^+$	241.0076
11	459.0253	$[2M + Na]^+$	$C_{14}H_{20}N_4NaO_4S_4^+$	459.0260

(B) Compound 7 UPLC/ESI$^+$-HRMSMS analysis. The enriched sample of compound **7** (see Appendix B, section 8) was dissolved in 500 µL of acetonitrile. 10 µL of the sample was injected into the UPLC/ESI$^+$-HRMS system described above, using a

water (0.1% formic acid)/acetonitrile (0.1% formic acid) gradient starting with 2 min at 5% acetonitrile, followed by a linear increase to 100% acetonitrile over 10 min, ending with 6 min at 100% acetonitrile. The MS was run in ESI^+ MSMS mode using a ramped collision energy of 15-45 eV. The range 50-500 *m*/z was scanned for daughter ions of $[M + H]^+$ = 392.1124 *m*/z.



Observed (*m*/z)	**Ion**	**Ion Formula**	**Calculated *m*/z**
392.1124	[M + H]+	$C_{18}H_{22}N_3O_3S_2^+$	392.1097
350.1017	$[M + H - C_2H_2O]^+$	$C_{16}H_{20}N_3O_2S_2^+$	350.0991
288.0497	$[M + H - C_8H_8]^+$	$C_{10}H_{14}N_3O_3S_2^+$	288.0471
229.0122	$[M - C_{10}H_{12}NO]^+$	$C_8H_9N_2O_2S_2^+$	229.0100
187.0006	$[M - C_{10}H_{12}NO - C_2H_2O]^+$	$C_6H_7N_2OS_2^+$	186.9994
171.9767	$[M - C_{10}H_{12}NO - C_2H_2O - CH_3]^{\cdot+}$	$C_5H_4N_2OS_2$	171.9760

(C) Targeted profiling of compounds in *S. clavuligerus* strains by MS.

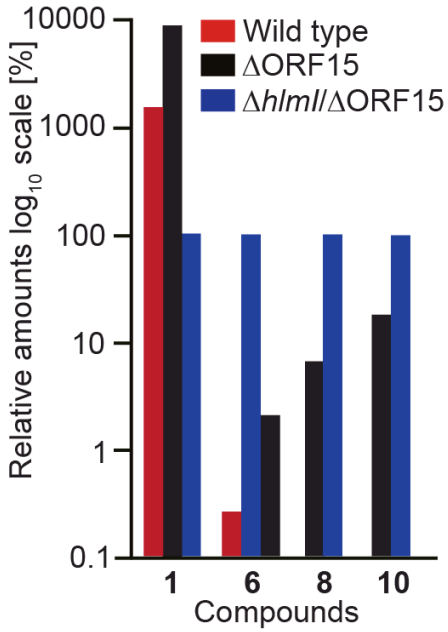


Figure B.1: Targeted compound profiling for strains: wild-type, ΔORF15, and Δ*hlmI*/ΔORF15. The amounts of compounds **1**, **6**, **8**, and **10** were determined relative to Δ*hlmI*/ΔORF15, arbitrarily set to 100 %. Relative values were measured using HPLC/ESI$^+$-MS ion chromatogram peak areas. Where no bar is present, the indicated metabolite was not detected. Compounds **3**, **5**, **9**, and **11**are omitted from this plot because they are only found in the Δ*hlmI*/ΔORF15 strain.

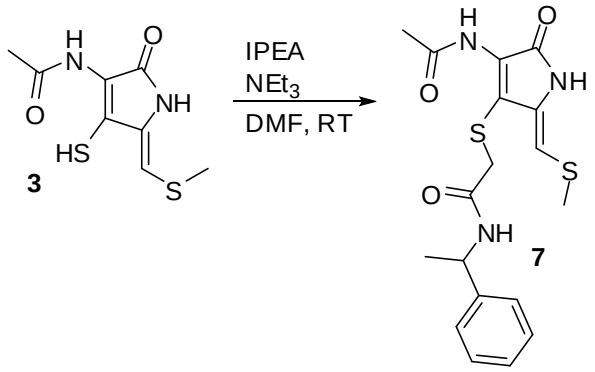
***8. Chemical Derivatization Procedures:* (A) *S*-alkylation of crude Δ*hlmI*/ΔORF15 supernatant extract with iodoacetamide.** A 10 mL culture of the Δ*hlmI*/ΔORF15 strain was grown at 30 °C for 4 days. A sample of 70 µL of the filtered culture supernatant was mixed with 10 µL of 300 mM iodoacetamide and 20 µL of 0.5 M NH_4HCO_3 buffer (pH 7.8). The reaction was incubated at room temperature for 30 min (protected from light) and quenched with 0.1% formic acid. A sample of 10 µL of the resulting mixture was injected into a LC/ESI$^+$-HRMS Q-Tof instrument (Agilent Technologies 6520) on a C18 column (Phenomenex Gemini-NX 5 µ, 50 × 2 mm) for detection of alkylated dihydroholomycin *via* positive electrospray ionization (ESI$^+$). The chromatography solvent consisted of a water (0.1% formic acid)/acetonitrile (0.1% formic acid) gradient starting with 2 min at 2% acetonitrile, followed by a linear increase to 100% acetonitrile

over 10 min, ending with 3 min at 100% acetonitrile. The ion chromatogram was generated using the mass spectral signal at *m*/z 288.0471 with 20 ppm error window.

Holomycin was isolated and purified from *S. clavuligerus* following the same procedure as previously described.[2] Alternatively, it was synthesized *via* the procedure of Hjelmgaard *et al.*[3] A sample of 5 µL of 1 mM holomycin, dissolved in DMSO, was mixed with 5 µL of 1.5 mM TCEP, 10 µL of 500 mM NH_4HCO_3 buffer (pH 7.8), and 25 µL of milli-Q water. A sample of 5 µL of 300 mM iodoacetamide was added to the reaction mixture and incubated at room temperature for 15 min. The reaction was quenched with 0.1% formic acid and filtered. A sample of 10 µL of the mixture was injected into the Q-Tof LC/ESI^+-HRMS instrument, running in ESI^+ mode, for detection of alkylated dihydroholomycin. Ion chromatograms were generated using the mass spectral signals at *m*/z 274.0315 (mono-IAA adduct) and 331.0529 (bis-IAA adduct, **4**) using a 20 ppm error window.

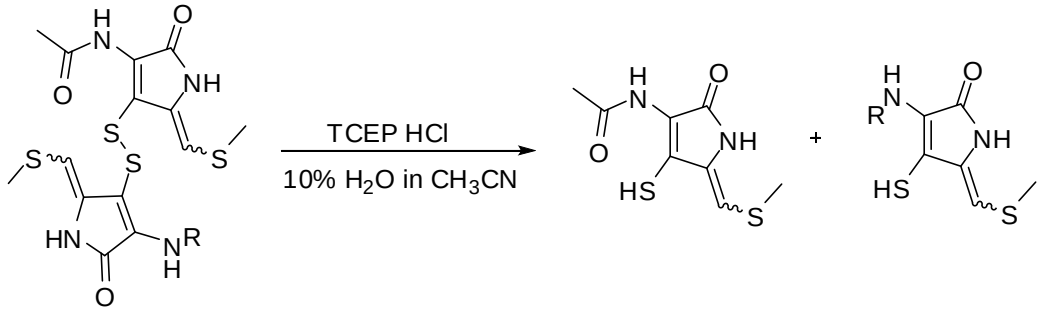
(B) Preparation of dimethyl-dihydroholomycin standards. Dimethyl-dihydroholomycin was prepared from holomycin under similar conditions as the diacetamidio-derivative (**4**) in a larger scale substituting iodoacetamide with methyl iodide. A mixture of isomers was generated and purified by HPLC. ^{1}H NMR (300 MHz; DMSO-d_6): δ(ppm) 10.22 (s), 9.39 (s), 6.17 (s), 2.46 (s), 2.32 (s), 1.97 (s). ^{13}C NMR (75 MHz; DMSO-d_6) δ(ppm) 169.3, 166.8, 134.2, 133.2, 129.4, 112.9, 55.8, 49.5, 23.4, 18.0, 16.8. ESI^+-MS *m*/z, calculated 245.0413, observed 245.0410 $[M + H]^+$.

(C) *S*-alkylation of enriched compound 3 with 2-iodo-*N*-(1-phenylethyl)acetamide. An enriched sample of **3** (Appendix B, section 4, pool 1 fractionation) was dissolved in 500 µL of dry DMF. While stirring, 50 µL of a DMF solution containing 2-iodo-*N*-(1-phenylethyl)acetamide (52 mM) and triethylamine (690 mM) was added. The reaction was stirred for 4 h, followed by removal of volatiles under rotary evaporation. The residue was redissolved in 0.3 mL of $CDCl_3$ and analyzed by NMR spectroscopy using a $CDCl_3$ susceptibility-matched Shigemi tube.



(D) TCEP reduction of disulfides in crude Δ*hlmI*/ΔORF15. HPLC/ESI$^+$-MS ion chromatogram of Δ*hlmI*/ΔORF15 extracts showed three chromatographic peaks with *m*/z values and isotope patterns consistent with the molecular ion of compound **3** (**Figure 2**), of which the earliest-eluting peak represented compound **3**, whereas the two late-eluting peaks additionally included mass spectral signals at *m*/z 417.0 and 439.0. These signals had diagnostic isotopic patterns suggesting the presence of four sulfur atoms. It was hypothesized that the signals at *m*/z 417.0 and 439.0 belong to a heterodimeric oxidation product of **3**, in which a disulfide bond has been formed between **3** and a non-*N*-acylated derivative of **3** (**9**, $[M + H]^+$ = 417.0 *m*/z, $[M + Na]^+$ = 439.0 *m*/z). A putative homodimeric product of **3** was also observed (**8**, $[M + H]^+$ = 459.0 *m*/z). The structures of **8** and **9** (**Figure 2**) were investigated using both chemical modification and UPLC/ESI$^+$-HRMS analysis.

To test the hypothesis that the disulfide homo (**8**) and heterodimers (**9**) stem from the oxidation of free **3**, a sample of Δ*hlmI*/ΔORF15 extract was reacted with the reducing reagent *tris*(2-carboxyethyl)phosphine (TCEP) and changes in the distribution of **3**, **8**, and **9** were monitored *via* HPLC/ESI$^+$-MS (see scheme below). 500 µL of a solution of TCEP (800 mM) in 10% water in acetonitrile (v/v) was added to 10 mg of Δ*hlmI*/ΔORF15 extract. 500 µl of 10% water in acetonitrile was added to 10 mg of a crude Δ*hlmI*/ΔORF15 extract as a negative control. Solutions were allowed to stir for 1 hour. The reaction mixtures were analyzed by HPLC/ESI$^+$-MS as described above in section 6, which showed that the Δ*hlmI*/ΔORF15 extract treated with TECP did not contain **8** and **9**, whereas the corresponding untreated Δ*hlmI*/ΔORF15 control produced peaks for both compounds. The HPLC/ESI$^+$-MS ion chromatogram of the TCEP-treated Δ*hlmI*/ΔORF15 extract contained four early eluting peaks with the mass and isotopic distribution of **3**; likely a combination of regio and stereoisomers of **3**. Follow-up UPLC/ESI$^+$-HRMS analysis of **8** and **9** showed exact masses and MS fragmentation supporting our structural hypotheses (section 7 above).



8, R = acetyl
9, R = H

9. UV Shifts of Holomycin and Derivatives:

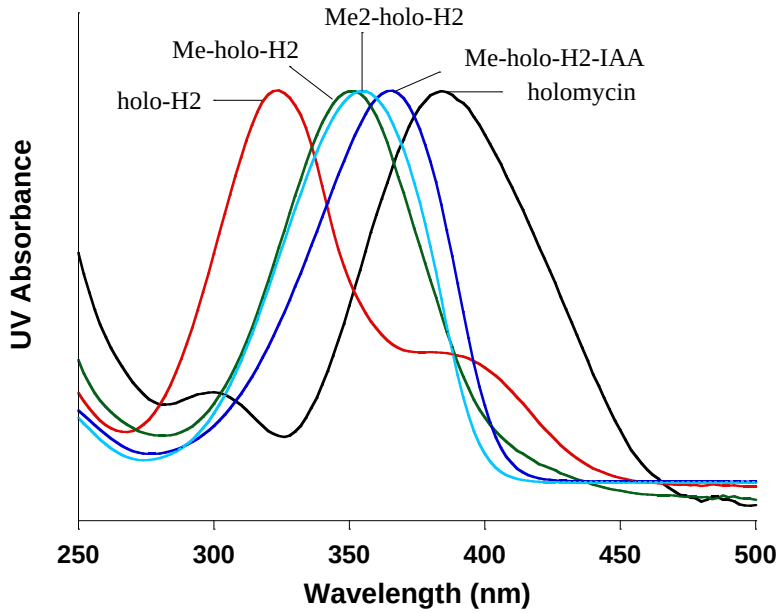


Figure B.2: UV spectra of holomycin and derivatives described in Chapter 2. Holo-H2, dihydroholomycin; Me-holo-H2, monomethyl-dihydroholomycin from the culture of ΔORF15/Δ*hlmI* mutant strain; Me2-holo-H2, synthesized dimethyl-dihydroholomycin standard; Me-holo-H2-IAA, iodoacetamide adduct of monomethyl-dihydroholomycin from the culture of ΔORF15/Δ*hlmI* mutant strain. The X axis is wavelength in nm and the Y axis is UV absorbance normalized to 100% for each spectrum.

10. Quantification of Compounds 3, 5, 6, 10 and 11 in Δhlmi/ΔORF15 Extracts: A 500 ml culture of Δ*hlmI*/ΔORF15 was extracted, fractionated into three pools, and analyzed by ^{1}H NMR spectroscopy as described above. To estimate the amounts of **3**, **5**, **6**, **10**, and **11**, the methine ^{1}H singlets, corresponding to H-1, were integrated relative to either the C<u>H</u>D_2CN (99.8%-*d*, d_3-acetonitrile) or C<u>H</u>D_2SOCD_3 (99.9%-*d*, d_6-DMSO) signals, depending on the NMR solvent used. The amounts of **3**, **5**, **6**, **10**, and **11** determined for each NMR sample were then divided by 500 mL to estimate the concentration of these metabolites in the Δ*hlmI*/ΔORF15 culture: **3**, 0.6 µM; **5**, 1.6 µM; **6**, 1.0 µM; **10**, 0.4 µM; and **11**, 0.3 µM.

11. Agar Diffusion Assay, Activity of Dimethyl-Dihydroholomycin: Spore stocks were prepared for wild-type, Δ*hlmI*, ΔORF15, and Δ*hlmI*/ΔORF15 strains. Approximately 5 × 10^5 spores were plated for each strain on MYM agar (4.0 g yeast extract, 10.0 g malt extract, 2.0 g dextrose, and 20.0 g agar per liter of media). Then a sample of 5 µL of 32.3 mM holomycin or dimethyl-dihydroholomycin dissolved in DMSO was added to the center of the plates. The plates were incubated at 30 °C for 4-7 days and the diameter of the inhibition zone was determined as a measurement for the sensitivity of these strains toward holomycin and dimethyl-dihydroholomycin. As observed previously, *hlmI* deletion strains (Δ*hlmI* and ΔORF15/Δ*hlmI*) were more sensitive to exogenously added holomycin than the wild-type and ΔORF15 strains, suggesting that the gene *hlmI* plays a role in self-protection in *S. clavuligerus*. In contrast, dimethyl-dihydroholomycin did not display any toxicity against all four strains tested. This result provides molecular evidence supporting the hypothesis that *S*-methylation is a self-protection strategy that *S. clavuligerus* utilizes in the absence of *hlmI*.

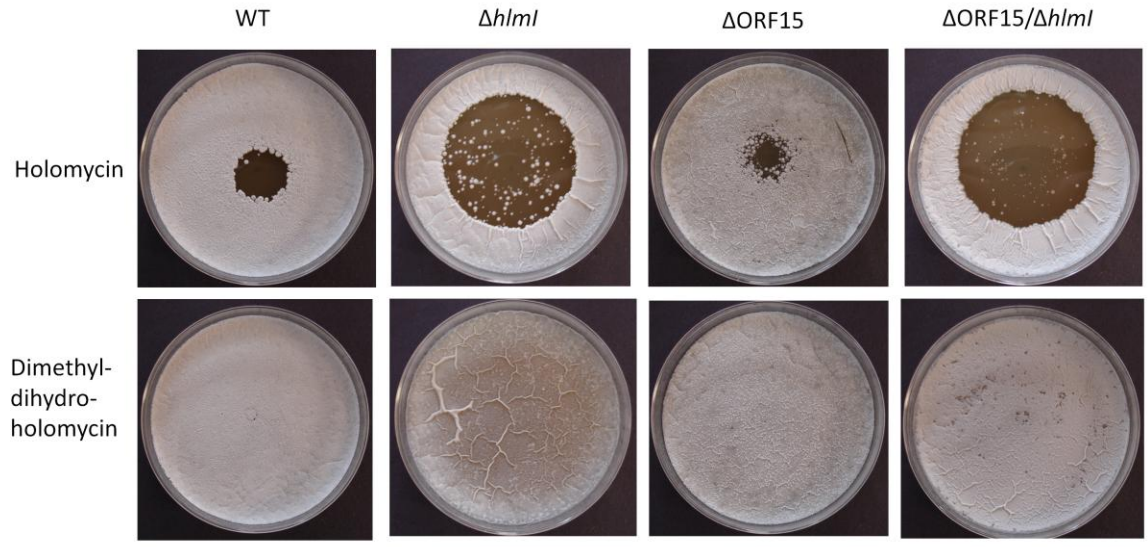
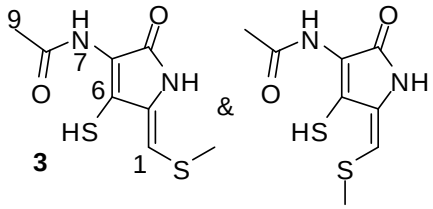


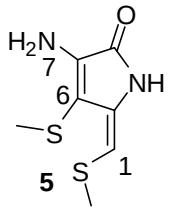
Figure B.3: Antimicrobial activities of dimethyl-dihydroholomycin and holomycin. Activity was compared amongst the *S. clavuligerus* wild-type (WT) and mutant strains.

Table B.1. Partial ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data, obtained in DMSO-d_6 for a Δ*hlmI*/ΔORF15-derived sample containing mixture of the *cis*- and *trans* stereoisomer of **3**. These data do not allow distinguishing between the *cis* and *trans* stereoisomers; therefore, the major metabolite is reported with the minor metabolite chemical shifts and correlations in parenthesis. Proton chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and ^{13}C chemical shifts to δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBC spectra. gHMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. Abbreviation: s = singlet.



Position	**δ_C**	**Proton**	**δ_H**	**gHMBC**
1	103.0 (101.9)	1-H	5.97 (s) (5.90) (s)	1-S-CH_3, 6, (1-S-CH_3, 6)
1-S-CH_3	16.7 (16.8)	1-S-CH_3	2.33 (s) (2.31) (s)	1, (1)
2				
3-N		3-NH		
4				
5				
6	165.1 (160.6)			
7-N		7-NH		
8	171.5 (171.1)			
9		9-CH_3	1.78 (s) (1.76) (s)	

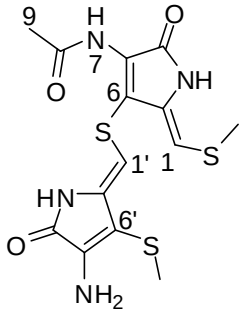
Table B.2. ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **5** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBC and HMQC spectra. gHMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. ROESY correlations were observed using a mixing time of 275 ms. Abbreviations: s = singlet, br. = broad.



5

Position	**δ_C**	**Proton**	**δ_H**	**HMBC**	**ROESY**
1	100.0	1-H	5.56 (s)	1-S-CH_3, 2, 6	1-S-CH_3, 3-NH, 6-SCH_3
1-S-CH_3	17.3	1-S-CH_3	2.33 (s)	1	
2	136.2				
3-N		3-NH	10.01 (s)	2, 5, 6	
4	165.2				
5	141.5				
6	99.2				
6-S-CH_3	17.1	6-S-CH_3	2.13 (s)	6, 6-S-CH_3	7-NH_2
7-N		7-NH_2	5.77 (br. s)	4, 6	

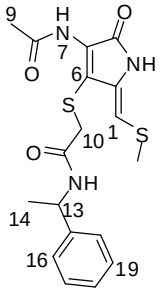
Table B.3. ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **6** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBC and HMQC spectra. gHMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. ROESY correlations were observed using a mixing time of 275 ms. Abbreviations: s = singlet, br. = broad.



6

Position	**δ_C**	**Proton**	**δ_H**	**gHMBC**	**ROESY**
1	112.7	1-H	6.18 (s)	1-S-CH_3, 2, 6	1-S-CH_3, 1', 3'-NH
1-S-CH_3	17.0	1-S-CH_3	2.40 (s)	1	6'-S-CH_3 (weak)
2	131.8				
3-N		3-NH	10.32 (s)	4, 5, 6	
4	168.9				
5	128.0				
6	129.5				
7-N		7-NH	not observed		
8	167.7				
9	22.3	9-CH_3	1.95 (s)	8	
1'	91.2	1'-H	5.57 (s)	6, 2'	6'-S-CH_3
2'	141.5				
3'-N		3'-NH	10.27 (s)	2', 4', 5', 6'	
4'	165.4				
5'	142.8				
6'	98.1				
6'-S-CH_3	17.2	6'-S-CH_3	2.11 (s)	6'	7'-NH_2
7'-N		7'-NH_2	6.13 (br. s)	4', 6'	

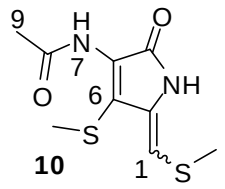
Table B.4. ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **7** in chloroform-*d* (and DMSO-d_6 where indicated). Chemical shifts were referenced to δ($C\underline{H}Cl_3$) = 7.26 and δ($^{13}\underline{C}HCl_3$) = 77.2 ppm. The gHMBC spectrum was referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined by (g)HMBC and HSQC spectra. (g)HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. *J* coupling constants were measured in the ^{1}H spectrum. Abbreviations: s = singlet, d = doublet, q = quartet, m = multiplet.



7

Position	δ_C	Proton	δ_H (J_{HH} [Hz])	(g)HMBC	ROESY
1	114.8	1-H	6.31 (s)	1-S-CH_3, 2, 6	10-H
1-S-CH_3	17.8	1-S-CH_3	2.49 (s)	1	
2	133.1				
3-N		3-NH	10.33(s), DMSO-d_6	5, 6, DMSO-d_6	
4	not observed				
5	129.5, DMSO-d_6				
6	125.7				
7-N		7-NH	not observed		
8	168.8				
9	23.8	9-CH_3	2.15 (s)	8	10-H
10	37.9	10-CH_2	3.51 (s)	6, 11	
11	166.7				
12		12-NH	not observed		
13	49.0	13-H	5.23 (dq, *J* = 7)	11, 15, 16, 20	
14	21.8	14CH_3	1.44 (d, *J* = 7)	13, 15	
15	142.7				
16, 20	125.9	16-H, 20-H	7.24 (d, *J* = 8)	18	
17,19	126.1	17-H, 19-H	7.33 (m)	15, 17, 19	
18	127.3	18-H	7.28 (m)		

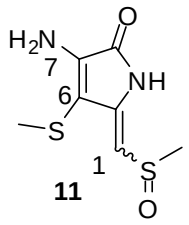
Table B.5. Partial ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **10** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBC spectra. gHMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. Abbreviation: s = singlet.



10

Position	**δ_C**	**Proton**	**δ_H**	**gHMBC**
1	111.7	1-H	6.18 (s)	1-S-CH_3, 2, 6
1-S-CH_3	16.6	1-S-CH_3	2.47 (s)	1
2	131.9			
3-N		3-NH	not observed	
4				
5				
6	133.0			
6-S-CH_3		6-S-CH_3	2.34 (s)	6
7-N		7-NH	not observed	
8	168.1			
9		9-CH_3	1.98 (s)	8

Table B.6. Partial ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **11** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm. ^{13}C chemical shifts were determined via gHMBC spectra. gHMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom. Abbreviation: s = singlet.



11

Position	**δ_C**	**Proton**	**δ_H**	**gHMBC**
1	105.4	1-H	5.90 (s)	1-S-CH_3, 2, 6
1-SO-CH_3	40.8	1-SO-CH_3	2.68 (s)	1
2	145.2			
3-N		3-NH	not observed	
4				
5				
6	95.6			
6-S-CH_3		6-S-CH_3	2.14 (s)	6
7-NH_2		7-NH_2		

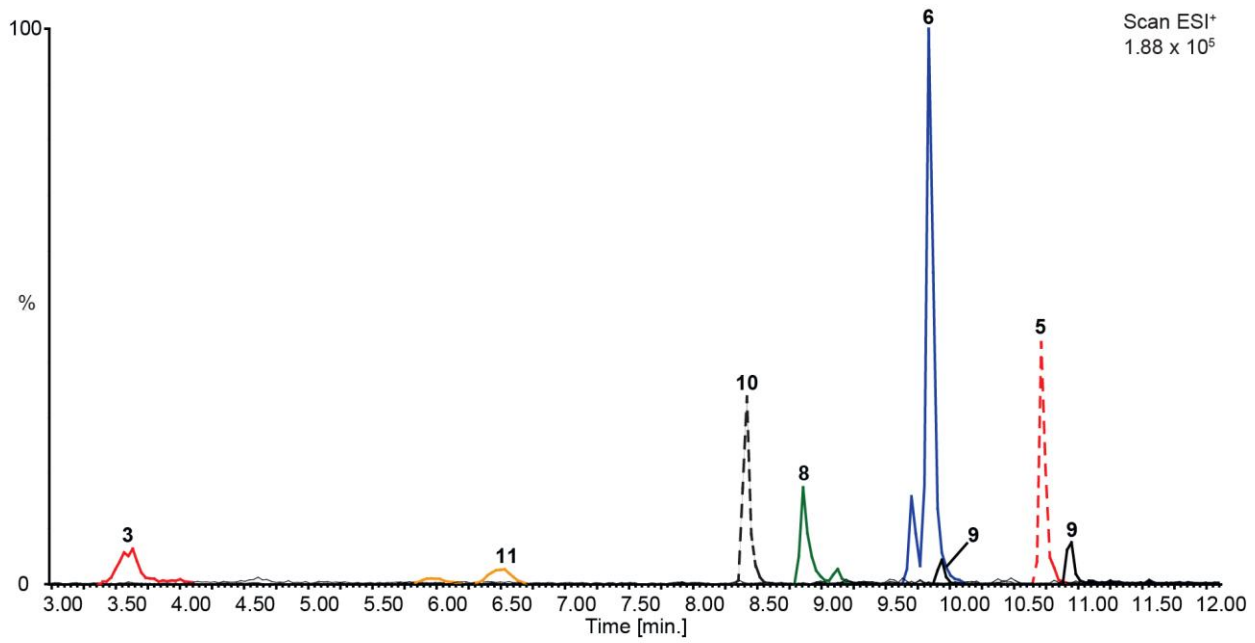


Figure B.4: Ion chromatograms representing $[M + H]^+$ for compounds 3, 5, 6, and 8-11. Chromatographic peak intensity was set relative to the that of the largest peak (compound **6**). The solid red trace corresponds to compounds **3**; yellow/orange trace, compound **11**; black dotted line, compound **10**; green trace, compound **8**; blue trace, compound **6**; solid black trace, compound **9**; and dotted red, compound **5**.

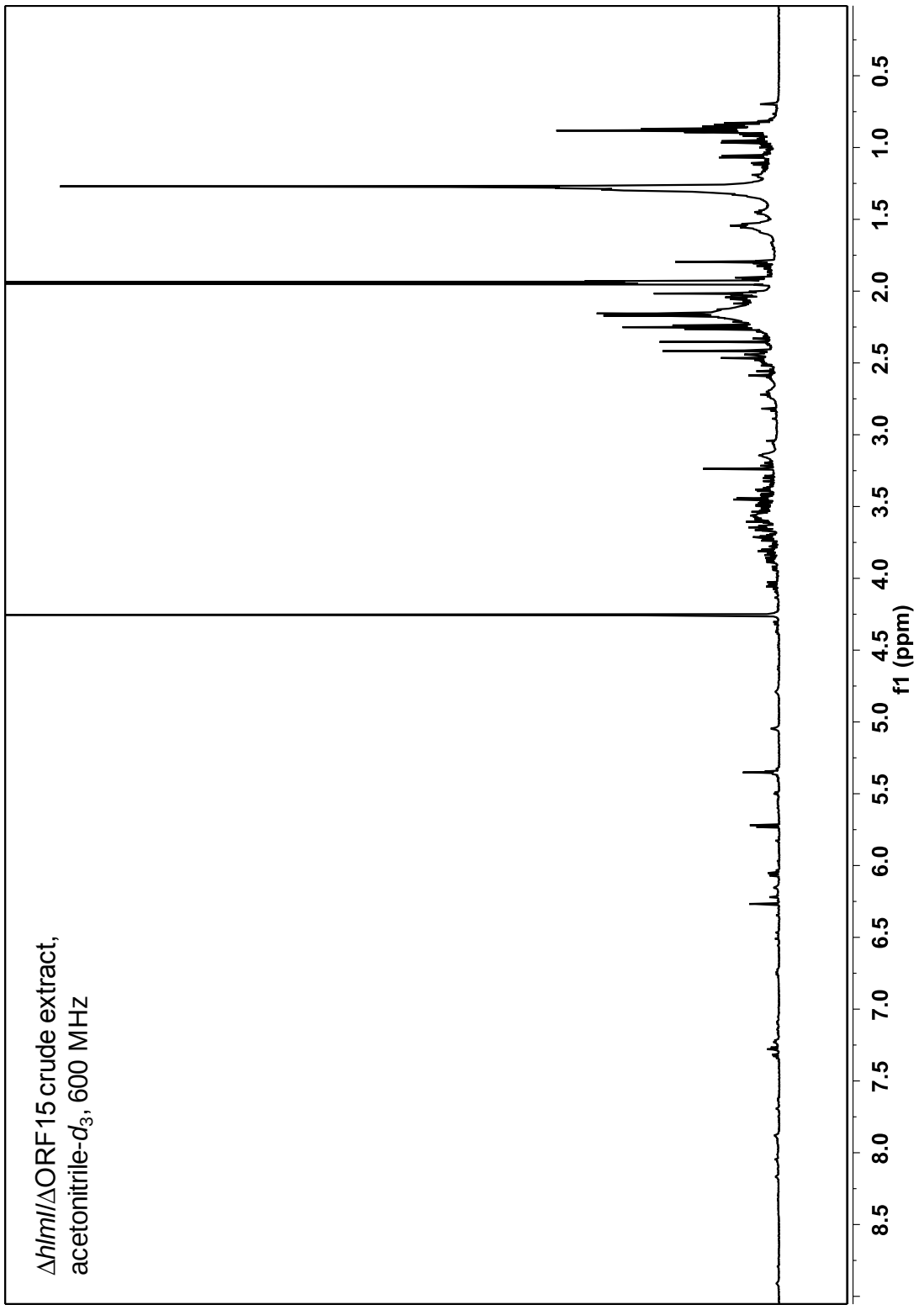


Figure B.5: ^{1}H NMR spectrum of crude Δ*hlml*/ΔORF15 extract, 600 MHz, acetonitrile-d_3.

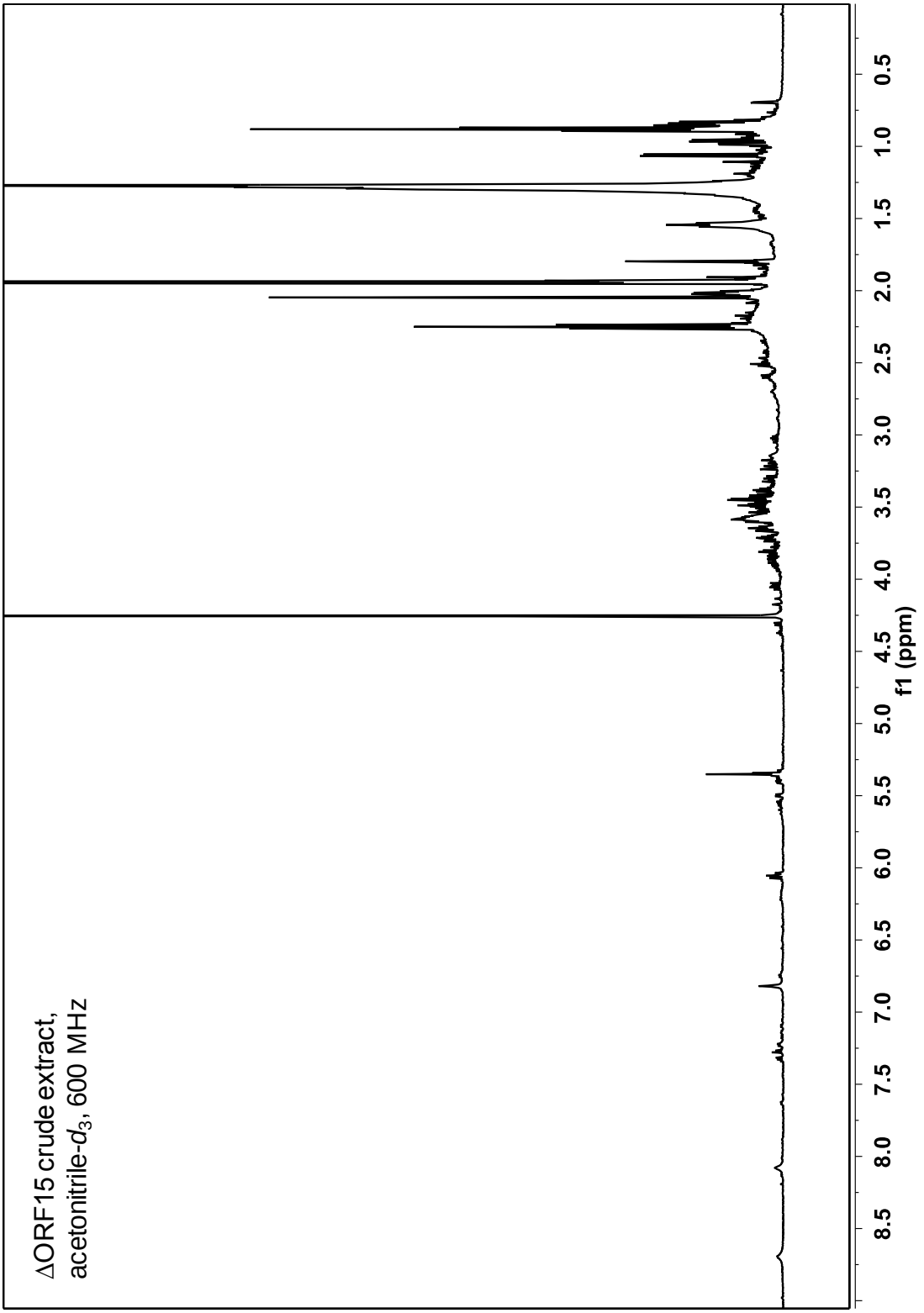


Figure B.6: ^{1}H NMR spectrum of crude ΔORF15 extract, 600 MHz, acetonitrile-d_3.

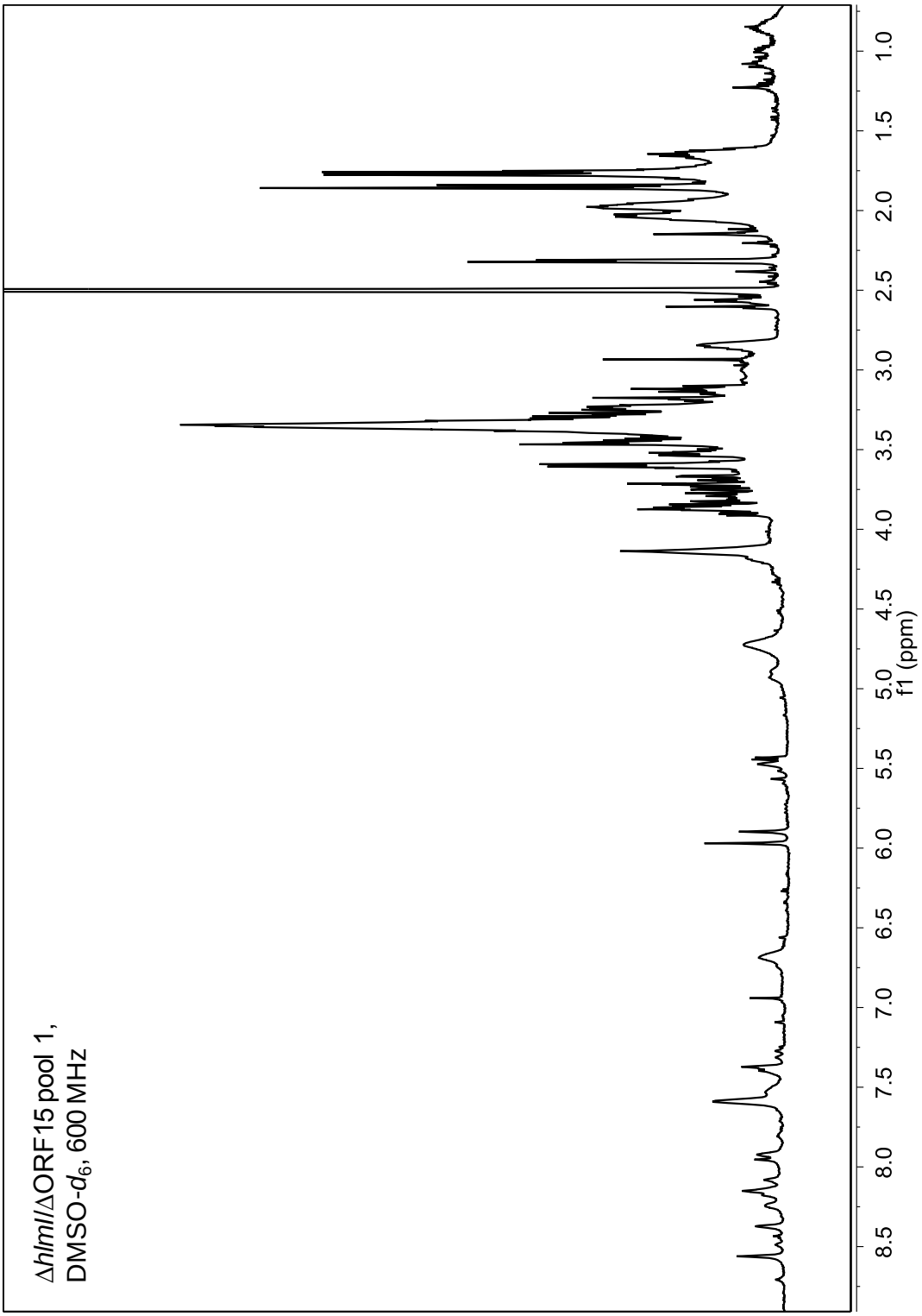


Figure B.7: 1H NMR spectrum of Δ*hlml*/ΔORF15 pool 1 compound **3**, 600 MHz, DMSO-d_6.

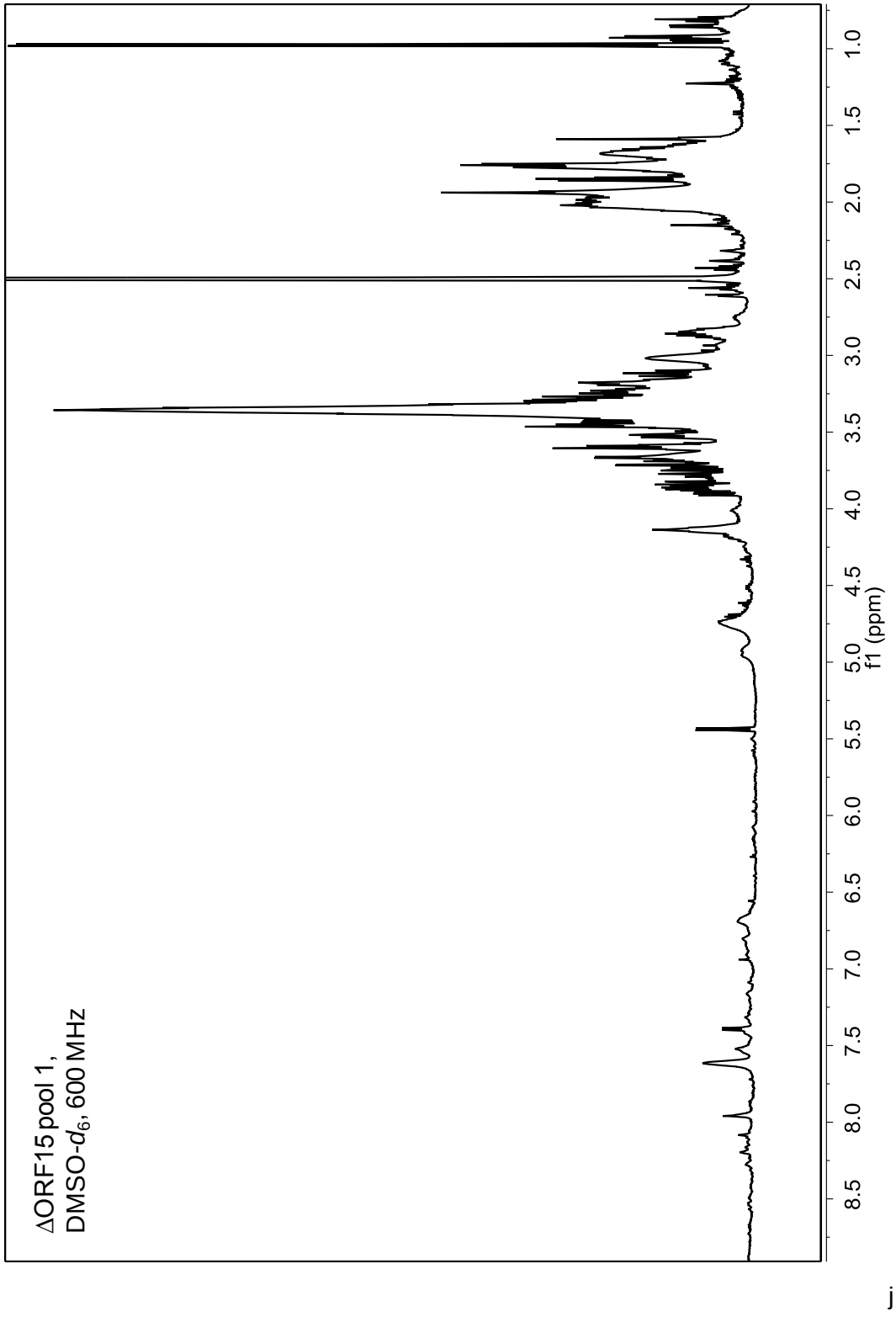


Figure B.8: ^{1}H NMR spectrum of ΔORF15 pool 1, 600 MHz, DMSO-$d_{6.}$

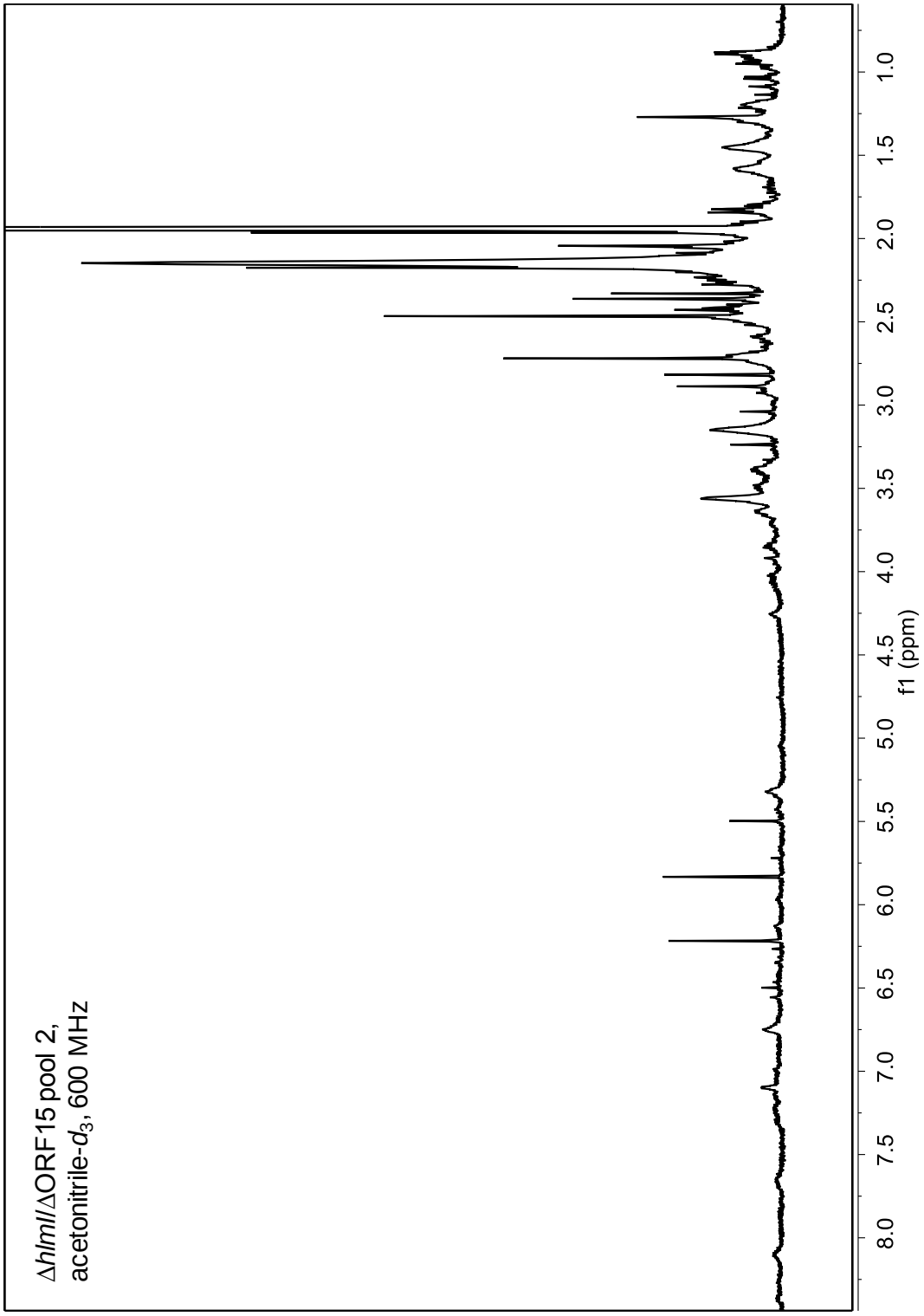


Figure B.9: ^{1}H NMR spectrum of Δ*hlml*/ΔORF15 pool 2 compounds **10** and **11**, 600 MHz, acetonitrile-d_3.

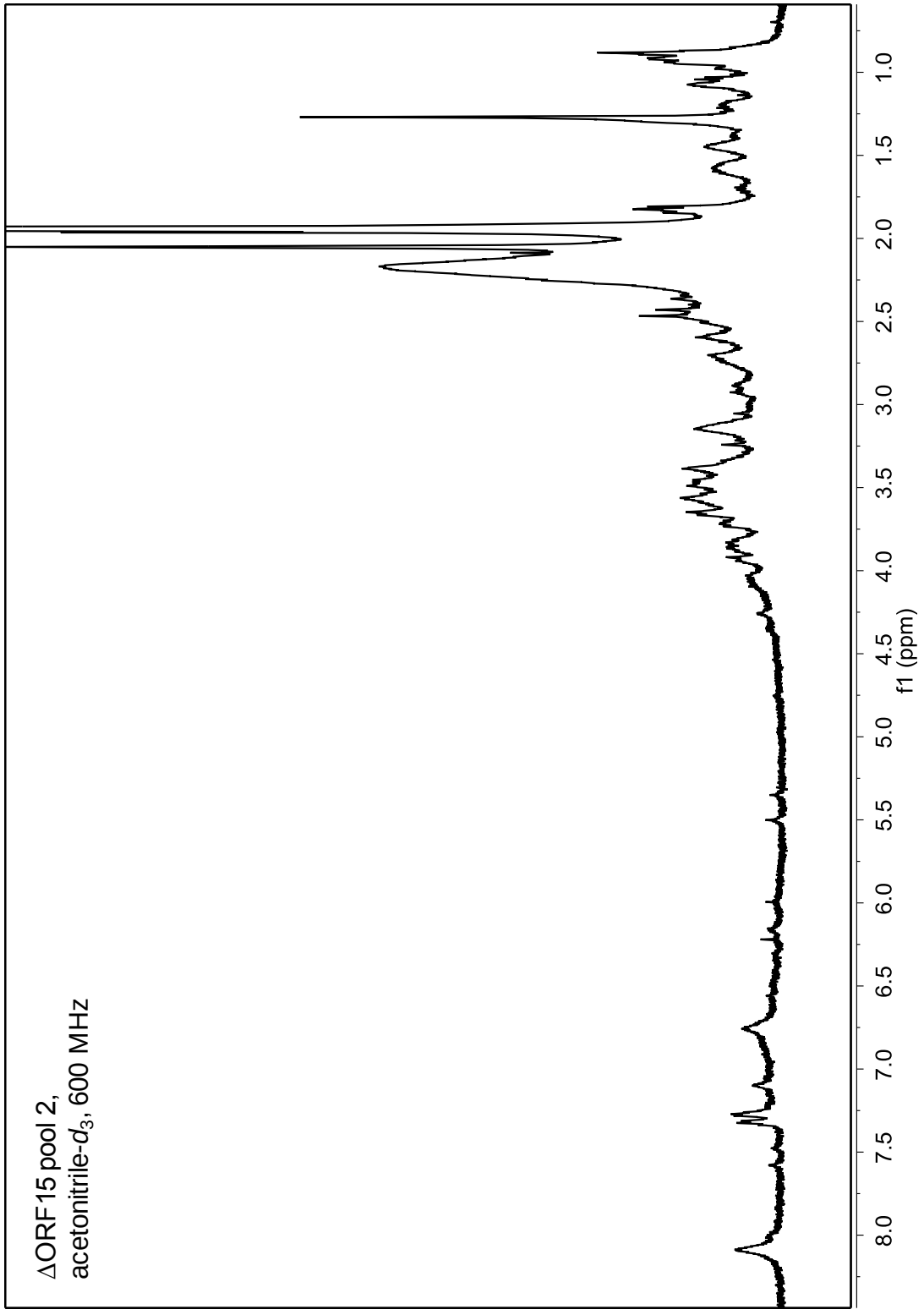


Figure B.10: ^{1}H NMR spectrum of ΔORF15 pool 2, 600 MHz, acetonitrile-d_3.

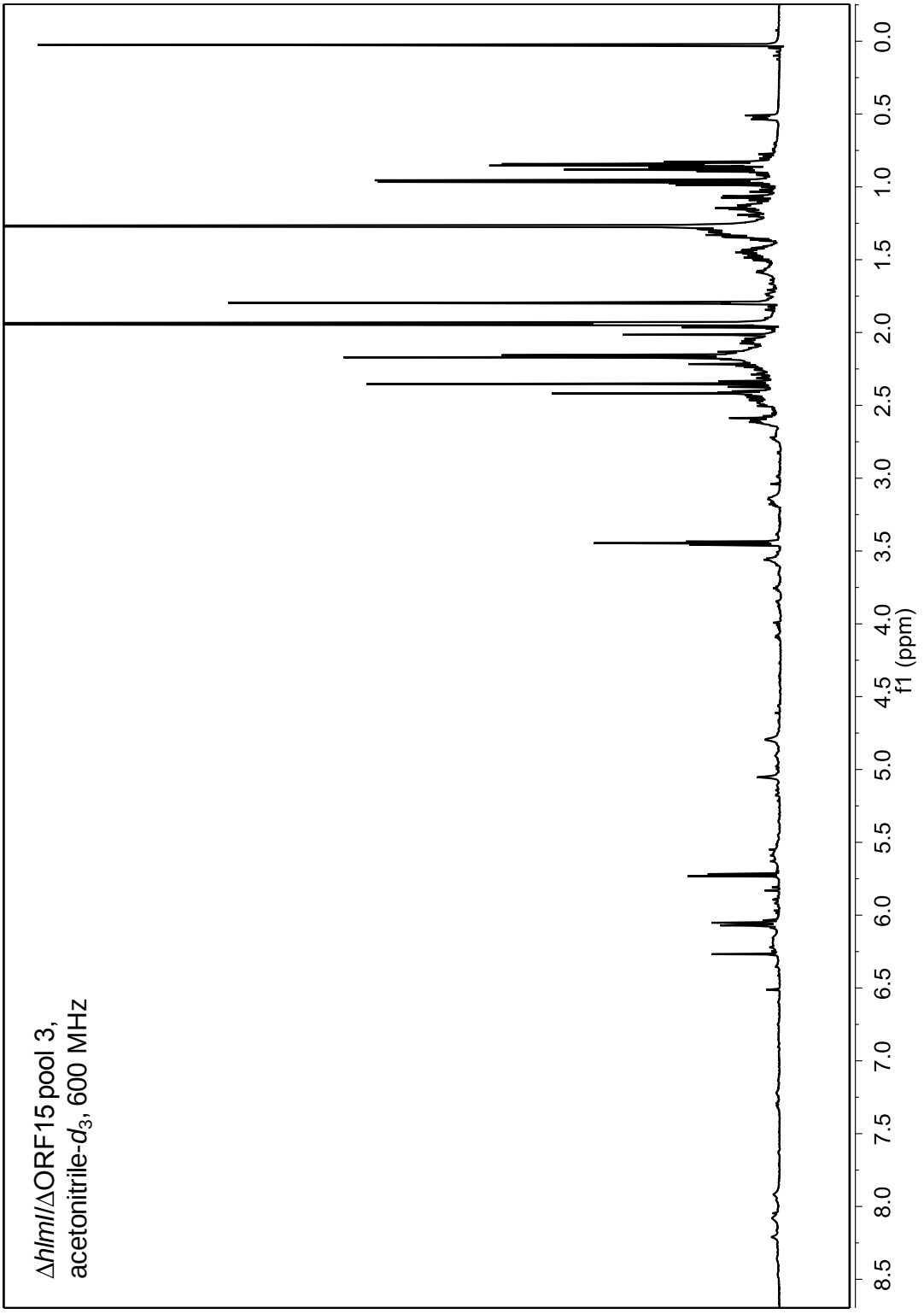


Figure B.11: ^{1}H NMR spectrum of Δ*hlml*/ΔORF15 pool 3 compounds **5** and **6**, 600 MHz, acetonitrile-d_3.

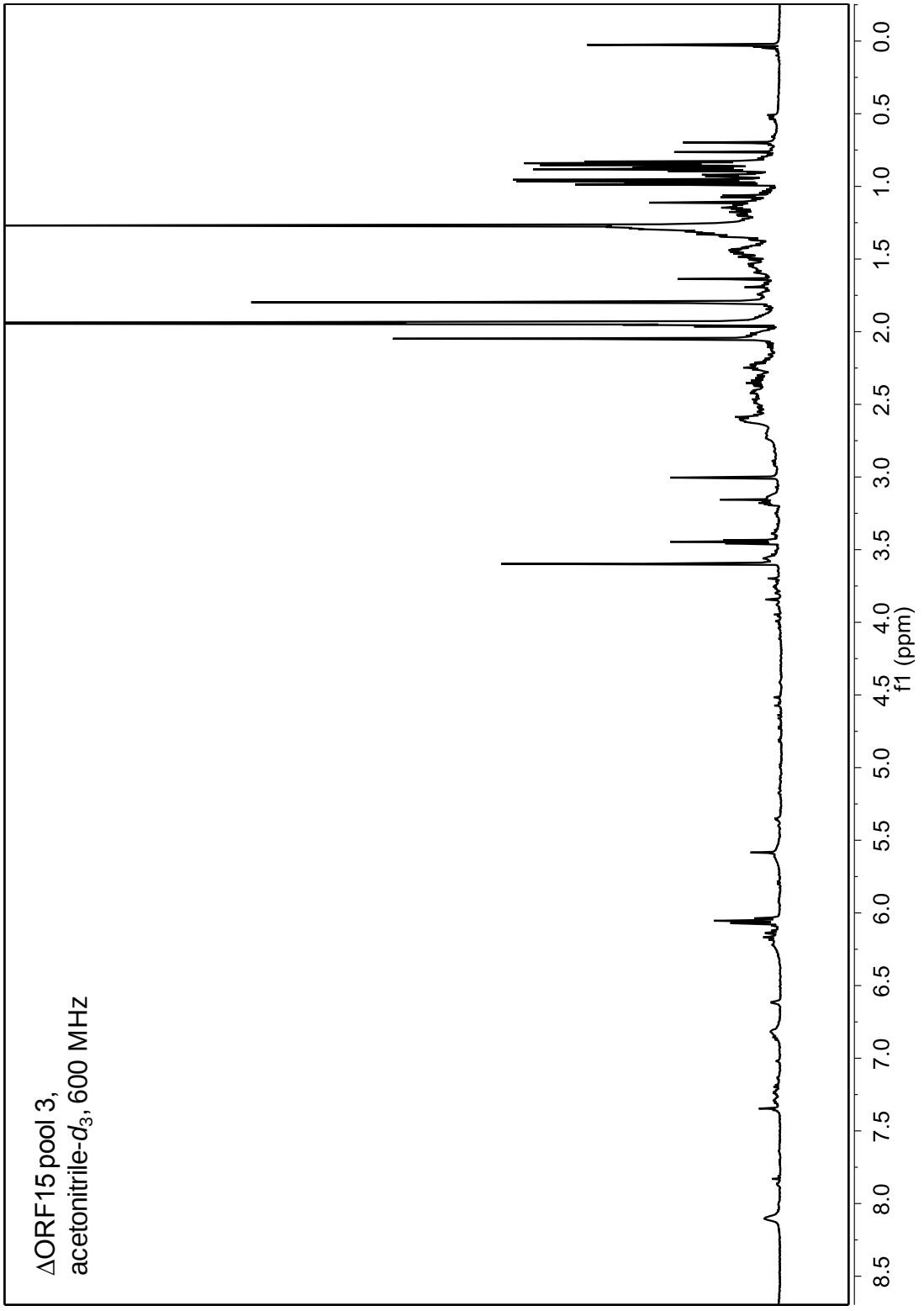


Figure B.12: ^{1}H NMR spectrum of ΔORF15 pool 3, 600 MHz, acetonitrile-d_3.

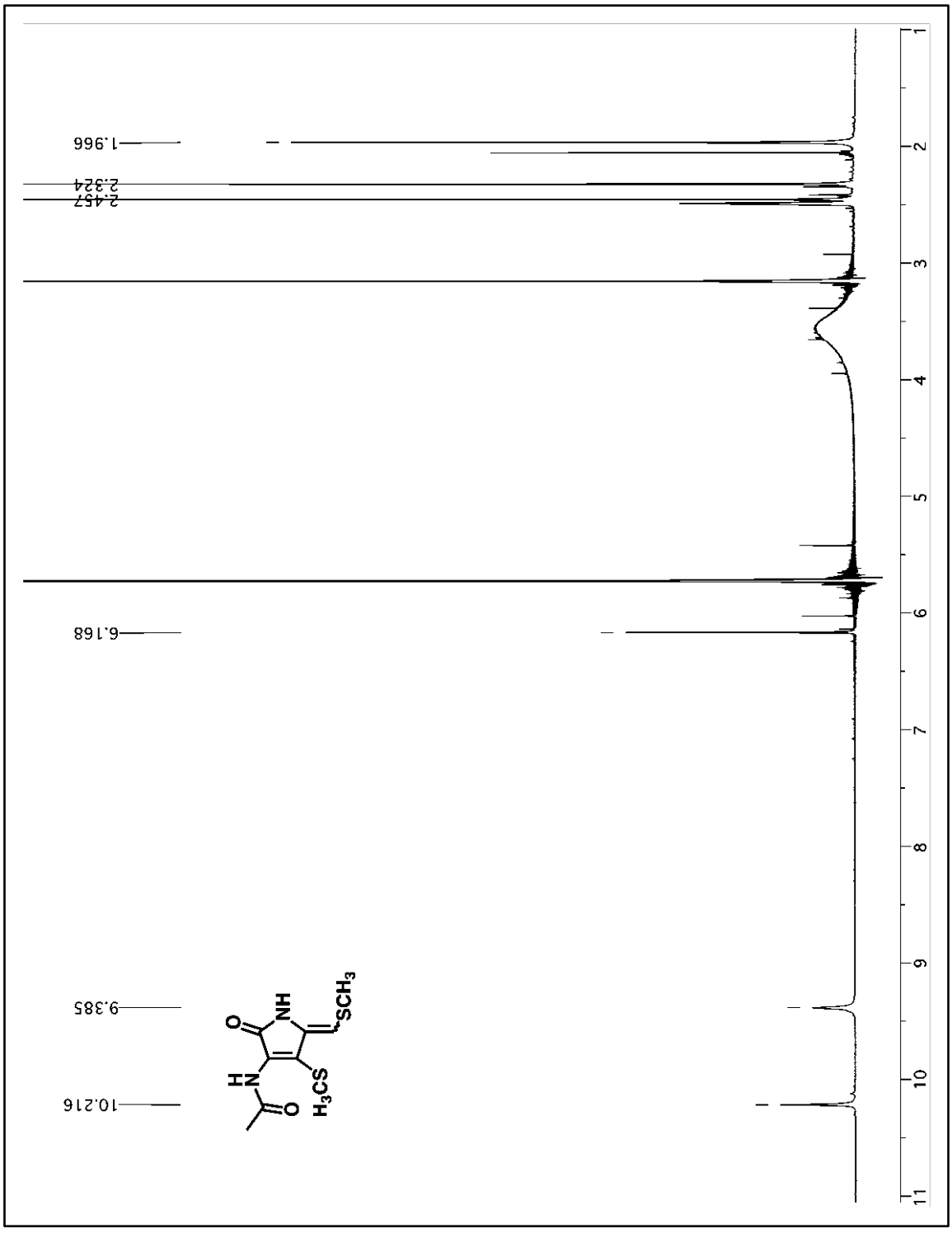


Figure B.13: ^{1}H NMR spectrum of synthetic dimethyl-dihydroholomycin, mixture of isomers, 300 MHz, DMSO-d_6.

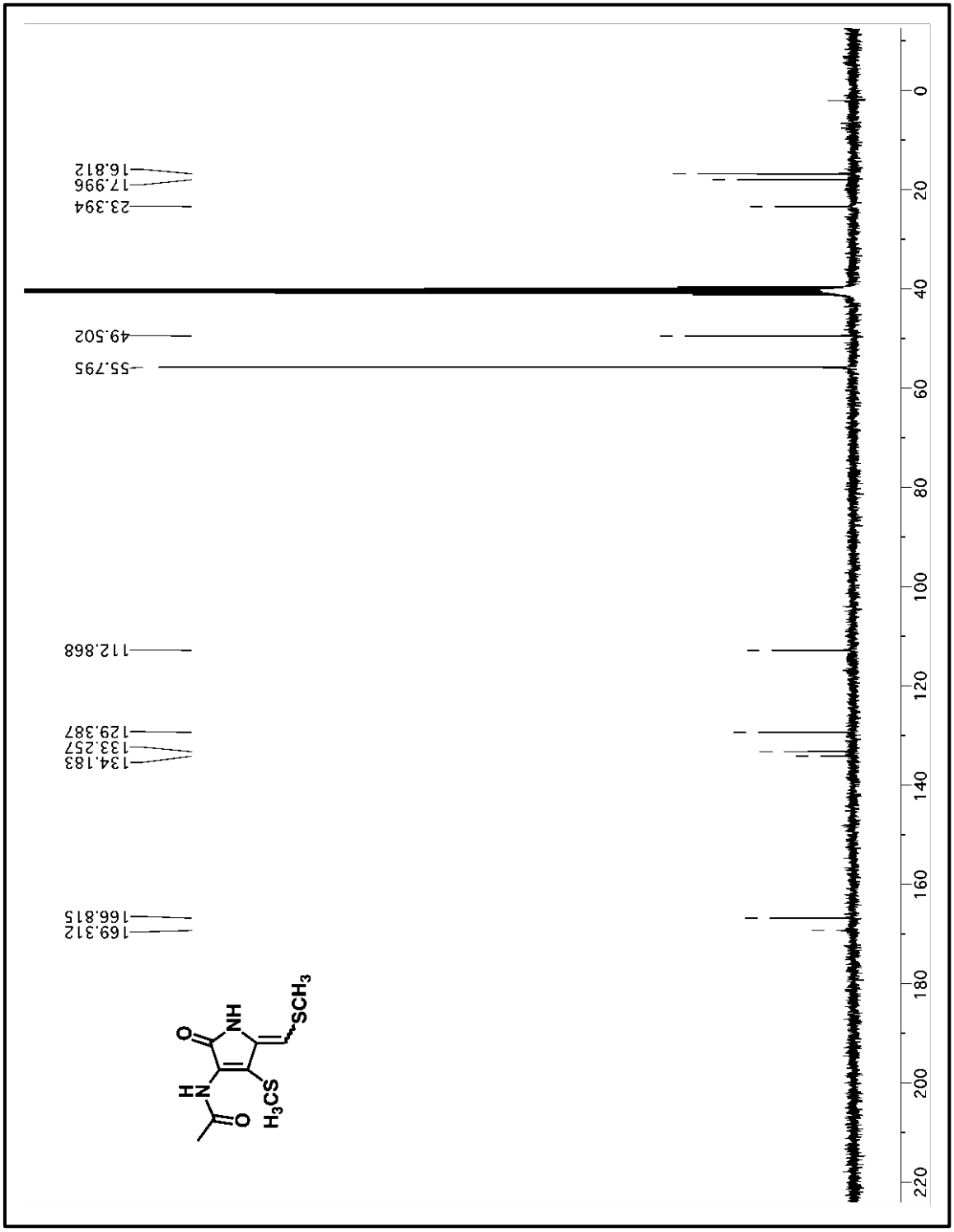


Figure B.14: ^{13}C NMR spectrum of synthetic dimethyl-dihydroholomycin, mixture of isomers, 300 MHz, DMSO-d_6.

REFERENCES

(1) Li, B.; Walsh, C. T. *Biochemistry* **2011**, *50*, 4615.

(2) Li, B.; Walsh, C. T. *PNAS USA* **2010**, *107*, 19731.

(3) Hjelmgaard, T.; Givskov, M.; Nielsen, J. *Org Biomol Chem* **2007**, *5*, 344.

Appendix C

HOMOLOGOUS NON-CANONICAL NRPS GENE CLUSTERS MEDIATE REDUNDANT SMALL-MOLECULE BIOSYNTHESIS IN *ASPERGILLUS FLAVUS*

1. *Construction of A. flavus Strains:* (A) Fusion PCR and Vector Construction. *lnaA* Constructs: Entire gene disruption and overexpression (OE) of *lnaA* constructs were created using fusion PCR.[1] All gene names are listed in **Table C.1**, all primers used in this study are listed in **Table C.2**, all strains are listed in **Table C.3**. For replacement cassettes, 1.5 kb upstream and downstream of *lnaA* open reading frame were amplified with primers DM101700-5'For/DM101700-5'Rev for upstream and DM101700-3'For/DM101700-3'Rev for downstream using NRRL3357 wild-type gDNA as template for PCR. These PCR fragments were fused with the *A. fumigatus* AF293 *pyrG* gene following.[2] Final PCR products were amplified with nested primers; DM101700-nestedFor and DM101700-nestedRev. The final PCR products with nested primers were confirmed by endonuclease digestion. To create the *OE::lnaA* cassettes, 4 PCR fragments were created and fused in this order: 1.5 kb of *lnaA* upstream fragment, 1.97 kb *A. fumigatus pyrG,* 1.5 kb *A. nidulans gpdA* promoter, and 1.5 kb *lnaA* open reading frame. First, the *A. fumigatus pyrG* and *A. nidulans gpdA* PCR fragments were fused using the same *A. fumigatus pyrG* primer set as described above and *gpdA*(p)*::pyrGFor* and *A. nidulans gpdA*(p)*Rev* primers for the *gdpA* fragment. Next, the *lnaA* upstream (using OE101700-5'For and OE101700-5'Rev) and *lnaA* open reading frame (OE101700-ORFFor and OE101700Rev) PCR fragments were fused to the *A. fumigatus pyrG*::*A.nidulans gpdA* promoter PCR fragment. The final construct was confirmed with restriction endonuclease and sequencing.

(B) Knockout and OE of *lnaA*. To create Δ*lnaA* and *OE::lnaA* mutants, protoplasts from freshly germinated conidia of *A. flavus* NRRL 3357.5 (*pyrG* auxotroph)

were used for transformation with a polyethylene glycol method.[3] Six µg of the correct PCR products were used for transformation. Transformants were identified by PCR using *A. fumigatus pyrG* primers for both the Δ*lnaA* and *OE::lnaA* strains. Further PCR was used to identify the Δ*lnaA* mutant (DM101700For and DM101700Rev) while *OE::lnaA* strains were confirmed using *A. nidulans* gpdAFor and Seq101700Rev3. Southern analysis confirmed the correct strains using a PCR generated probe with DM101700nested For and Rev, **Figure C.8**.

(C) *lnbA* Silencing. An RNAi silencing construct for *lnbA* from cluster 48 was generated in the following manner. First, a 528 bp PCR fragment of *lnbA* was amplified from wild-type gDNA using IRT121520 5FFor and 5FRev, and this was then inserted into the NcoI and AscI site of pTMH44.1[4] through quick-change method[5] to create pSA20.3. After confirming the sequence and direction of the *lnbA* insert using PCR with primers with pTMH44.1seq1For and seq1Rev, the same PCR fragment was amplified with IRT121520 3FFor and Rev from wild-type gDNA and ligated into the NotI and BamHI site of pSA20.3 to create pSA21.7. The vector was confirmed by PCR with primers pTMH44seq1For and seq2Rev, endonuclease digestion, and sequencing using pTMH44seq2For and seq2Rev. For construction of the TSA 22.1 or TSA 22.6 (Δ*lnaA*, *KD::lnbA*) strains were surveyed for stable genetic incorporation of *KD::lnbA* in the Δ*lnaA* background *via* Southern analysis, see **Figure C.9**

(D) *lnaB* Silencing. A RNAi silencing construct for the *nmrA*-like gene (AFLA_101710), *lnaB,* was generated in the following manner. First, a 579 bp PCR fragment was amplified from wild-type gDNA using IRT101710 5F For and 5F Rev, and inserted into the NcoI and AscI site of pTMH44.1[4] through quick-change method[5] to create pSA6.1. After confirming the sequence and direction of the *lnaB* insert using PCR with primers with pTMH44seq1For and seq1Rev, the same PCR fragment was amplified with IRT101710 3F For and Rev from wild-type gDNA and ligated into the *NotI* and *Hind*III site of pSA6.1 in the opposite direction to create pSA7.3. The vector was

confirmed by PCR with primers pTMH44seq1For and seq2Rev, endonuclease digestion, and sequencing using pTMH44seq2 For and seq2Rev.

(E) Northern Analysis *lna* Gene Cluster Expression. Total RNA was extracted from fungal cultures using the Trizol method. Blots were hybridized with DNA fragments amplified from NRRL3357 gDNA using the primers of all cluster genes: *lnaA*, *lnaB*, *lnaC*, *lnaD*, *lnaE*, *lnaF*, as well as *laeA* as listed in **Table C.2**. Detection of signals was carried out with a Phosphorimager-SI (Molecular Dynamics).

2. Fungal Strains and Experimental Growth Conditions: **(A) Fungal Strains**. NRRL 3357 (wild-type), NRRL 3357.5 (*pyrG*), TSA5.21 (Δ*lnaA*), TSA 7.15 (*OE::lnaA*), TSA 7.51 (*OE::lnaA*), TSA 22.1 (Δ*lnaA*, *KD::lnbA*), TSA 22.6 (Δ*lnaA*, *KD::lnbA*) and TSA16.7.31 (Δ*lnaA*, *KD::lnaB*). See **Table C.3** for corresponding genotypes.

(B) Experimental Growth Conditions. All strains were maintained as glycerol stocks. To setup minimal media experimental cultures, approximately 2 wk old *A. flavus* cultures grown on solid glucose minimal media plates (GMM [6]), amended with uridine [1.26 g/L], uracil [0.56 g/L], sorbitol [20 g/L], and for solid culture Difco Noble agar [15 g/L]) were flooded with 1-5 mL of a 0.1 % Tween 20 solution. A sterile cell scraper was used to remove conidia from the mycelia, and the resulting spore suspension was diluted 100 fold. Conidia concentration were measured for the 100 fold diluted conidia suspensions using a hemocytometer, and appropriate dilutions, using the aqueous 0.1% Tween 20 solution, were made to achieve desired conidia concentrations. Two methods were used to grow experimental GMM cultures (1) liquid stationary cultures: appropriate aliquots of conidia suspension were used to inoculate 6-cm Petri plates containing 11 mL of liquid GMM (initial plate conidia concentration 10^3 -10^7 spores / plate) (2) molten spore suspension: Petri plates containing 8 mL of solid GMM were overlaid with 3 mL of a molten solid GMM spore suspension (Difco Noble agar 7.5 g/L). Cultures were incubated in the dark at 29 °C for seven days.

3. Metabolite Extraction: To ensure chemical comparability among a given set of *A. flavus* strains, cultures were grown at the same time, on the same batch of media, and extracted using the same solvent mixtures and extraction protocols. **(A) Generation of initial crude extracts for DANS**. Solid rich media cultures of Δ*lnaA*, wild-type, and *OE::lnaA*, consisting of three plates each, were cultured as described above. Whole cultures were cut up into ~ 1 cm^2 pieces. The cultures were transferred into individual round bottom flasks, frozen (-80 °C), and lyophilized. Cultures were then extracted with 100 mL of ethyl acetate and filtered over ethyl acetate-washed cotton. The extraction process was repeated twice more with 25 mL of ethyl acetate. Solvent was removed from the extract solutions using rotary evaporation at room temperature.

(B) Extraction of Additional Rich Media- and GMM-Cultured Fungal Samples. Whole liquid stationary cultures, media and cellular material, not immediately processed were frozen and stored at -80 °C. Subsequent processing of liquid cultures entailed thawing the frozen cultures, transferring the samples to glass Erlenmeyer flasks, freezing the samples (-80 °C), and lyophilizing. Solid plate cultures were processed up to the extraction step as described above. The lyophilized material from solid or liquid cultures was extracted using 10% methanol in ethyl acetate (10-20 mL solvent / plate in sample), and filtered over methanol and ethyl acetate-washed cotton. The extraction was repeated twice more using 2.5-5.0 mL solvent / plate. Solvent was removed as described above.

4. Analytical Methods and Equipment Overview: **(A) NMR spectroscopy**. NMR spectrometers: Varian INOVA 600 MHz NMR spectrometer (600 MHz ^{1}H reference frequency, 151 MHz for ^{13}C), equipped with an HCN indirect-detection probe, and a Bruker AVANCE 600 MHz NMR spectrometer (601 MHz ^{1}H reference frequency, 151 MHz for ^{13}C) equipped with a 5 mm TXI probe. Non-gradient phase-cycled dqfCOSY spectra were acquired using the parameters: 0.6 s acquisition time, 500-600 complex

increments, 8-32 scans per increment. ROESY spectra were acquired using the parameters: 0.25 s acquisition time, 0.2 s mixing time, 400 increments, 16 scans per increment. Gradient and non-gradient HSQC[AD], HMQC, and HMBC[AD] spectra were acquired with these parameters: 0.25 s acquisition time, 300-600 increments, 4-32 scans per increment. $^{1}H,^{13}C$-HMBC spectra were optimized for $J_{H,C}$ = 6 Hz. Susceptibility-matched NMR tubes (Shigemi) were used for sample amounts smaller than 2 mg. NMR spectra were processed and baseline corrected using Varian VNMRJ and MestreLabs MestReC and MNOVA software packages.

(B) Mass spectrometry. High-resolution mass spectrometry was performed on a Waters nanoACQUITY UPLC system equipped with a Waters Acquity UPLC HSS C18 column (2.1 x 100 mm, 1.8 µm particle diameter) connected to a Xevo G2 QTof Mass Spectrometer operated in either ESI^+ or ESI^- modes, or an LTQ Orbitrap Velos (Thermo Scientific) mass spectrometer operated in ESI^+. HPLC/ESI^+-MS data acquisition and processing was controlled by Waters MassLynx software.

(C) Chromatography. Flash chromatography was performed using a Teledyne ISCO CombiFlash system. For semi-preparative HPLC a Supelco Discovery HS C18 column (25 cm X 10 mm, 5 µm particle diameter) was employed. An Agilent Zorbax Eclipse XDB-C8 column (4.6 X 150 mm, 5 µm particle diameter) was used in the HPLC/ESI^+-MS *A. flavus* mutant profiling analysis.

5. Generation of OE::lnaA(TSA7.51) and ΔlnaA Pools 1-3 for DANS Comparison: Twenty rich media cultures of *OE::lnaA* and Δ*lnaA* were cultured and extracted as described above (section 2). Extracts were loaded onto a silica gel loading column using dichloromethane. Residual material was dissolved in 50/50 methanol/DCM solution (v/v) and pipetted onto the loading column. The sample was then flash chromatographed using a CombiFlash equipped with a 40 g silica gel Gold column. A DCM-methanol gradient was used (40 mL/min) starting with 0% methanol for 6 minutes,

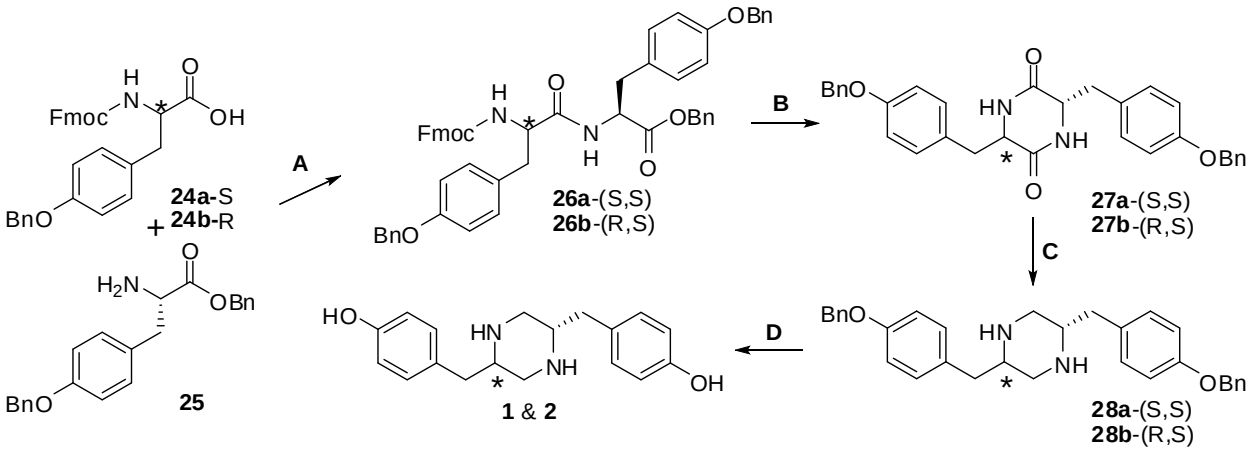
followed by a linear increase to 10% methanol over 12 minutes, followed by a linear increase to 30% methanol over 12 minutes and holding at 30% methanol for an additional 6 minutes. Fractions were pooled, generating the least polar pool 1, fractions 1-20 (0-2% methanol); pool 2, fractions 21-40 (2-10 % methanol); and the most polar pool 3, factions 40-76 (10-30 % methanol).

6. Differential Analysis by 2D NMR Spectroscopy: High-resolution dqfCOSY spectra were acquired as described above. Spectra were zero-filled to 8192 complex data points in the directly detected dimension (F2) and 4096 data points in the indirectly detected dimension (F1). Bitmaps derived from absolute-value processed dqfCOSY spectra were then imported into Adobe Photoshop CS3 and overlayed as described in reference[7].

7. Purification and Characterization of Compounds 3-8: To setup rich media cultures, approximately 2 wk old *A. flavus* cultures grown on solid glucose minimal media plates (see section 2) were flooded with 1 mL of an aqueous 0.1% Tween 20 solution. After light agitation, 7 µL of the resulting spore suspension was point-inoculated onto solid Bacto tryptic soy media (30 g/L, rich media) plates (9-cm Petri dishes with 23 mL media/plate), and incubated at 25 °C in the dark for seven days. Extracts derived from 120 rich media-grown *OE::lnaA* (TSA7.51) cultures (media and mycelia) were divided into four portions and filtered through a RediSep silica gel column. Fractions were pooled, and solvent was removed using rotary evaporation at room temperature. The extract was then dissolved in methanol and adsorbed onto 1 g of octadecyl-functionalized silica gel. After removal of solvent under rotary evaporation, the sample was placed into a loading column. The sample was chromatographed using a Teledyne ISCO CombiFlash system equipped with a 100 g RediSep High Performance Gold C18 reverse phase column. The sample was chromatographed using

a 0.03% acetic acid in water and 10% acetonitrile in methanol ("organic") as solvents, using a gradient starting with 0% organic for 15 minutes, followed by a linear increase to 25% organic over 21 minutes, followed by a linear increase to 50% organic over 9 minutes, followed by a linear increase to 80% organic over 8 minutes, followed by a linear increase to 100% organic over 3 minutes, and holding at 100% organic for 15 minutes. Fractions were evaporated to dryness individually using rotary evaporation at room temperature. Every other fraction in the polarity range 18–54% organic (fractions 59–97) was dissolved in methanol-d_4 followed by acquisition of ^{1}H and dqfCOSY. Fractions containing **3**-**8** were combined as appropriate and used to acquire additional dqfCOSY, (g)HMBC, and HMQC[AD] or HSQC[AD] spectra (see **Tables C.4**-**C.11**).

8. Synthesis of Compounds 1 and 2:



Scheme C.1: Total synthesis of compounds 1 and 2. The synthetic route to **1** (* = S configuration) starting with **24a**; **2** (* = R stereo configuration) starting with **24b**. Reagents and conditions: (**A**) PyBOP, HOBt, DIEA CH_3CN, room temperature (rt), yield: **26a**, 60%; **26b**, 25%. (**B**) 6 eq. piperidine in DMF, rt, yield: **27a/b**, 50%. (**C**) LAH, Et_2O, 60 °C; yield: **28a**, 25%; **28b**, 77%. (**D**) H_2(g), Pd/C in EtOH w/ and acetic acid, rt, yields: **1**, 90%; **2**, 95%.

The four-step total synthesis of compounds **1** and **2** is depicted in **Scheme C.1**. In all cases product formation was confirmed by ^{1}H NMR. In step **A**, (*S*)-tyrosine benzyl ester (**25**) was coupled with Fmoc-protected (*R*)- or (*S*)-tyrosine (**24a/b**) following the procedures used by Campo *et al.* in their synthesis of a similar diketopiperazine cyclo(Phe-Ser)[8]. In step **B**, **20a** or **20b** was cyclized to the corresponding diketopiperazines **27a** or **27b** respectively, using the procedures described by[8]. Products were characterized using ^{1}H NMR, dqfCOSY, ^{1}H,^{13}C-HMQC, and ^{1}H,^{13}C-HMBC spectra.

Reduction and debenzylation. In step **C**, samples of the diketopiperazines of **27a** or **27b** (80 mg, 0.16 mmol) were reduced to the corresponding piperazines using LAH (60 mg, 1.6 mmol) in dry ether (5 mL). The reaction was stirred in a sealed glass vial at 60 ˚C for 1 hour. After cooling to room temperature, solvents were removed *in vacuo*. To the residue, 2 N aqueous HCl (6 mL) was added and the mixture was stirred for 10 minutes. After acidification, saturated Na_2CO_3 solution was added until basic. The mixture was extracted with 10% methanol in dichloromethane (4 x10 mL). The combined organic extracts were dried over sodium sulfate. The extract was adsorbed onto silica *in vacuo* and flash chromatographed using a DCM/methanol gradient. Products were obtained in 25% (**28a**) and 77% (**28b**) yield. In step **D**, solutions of **28a** or **28b** (5 mg, 0.010 mmol) in ethanol (15 mL) containing ~50 µL acetic acid were hydrogenated using H_2 gas and 10% palladium on activated carbon (Pd/C, 2 mg). Starting material, solvent, and Pd/C were degassed under argon, followed by the introduction of H_2 gas at atmospheric pressure. The reaction was stirred for 12 hours and then filtered through ethanol-washed cotton using 100 mL of ethanol containing ~200 µL acetic acid. Purification was carried out by loading the filtrate onto Celite, followed by flash chromatography using a RediSep High Performance Gold C18 reverse phase column. Compounds **1** and **2** were obtained in 90% and 95% yields respectively. HPLC retention times, and NMR spectra

of the products matched those of natural **1** and **2** derived from *A. flavus* extracts (For NMR spectroscopic data of compounds **1**, **2**, **27a**, and **27b** see **Tables C.4**, **C.5**, **C.12**, and **C.13** respectively).

9. Chromatographic Enrichment Compound 10: Wild-type and *OE::lnaA* fractions containing **10** were combined in methanol, silica gel was added, and the solvent was removed using rotary evaporation. The resulting silica gel-loaded sample was flash chromatographed using the Teledyne ISCO CombiFlash system equipped with a 24 g RediSep silica gel column. A dichloromethane/methanol gradient was used (35 mL/min) starting with 3 minutes at 0% methanol, followed by a linear increase to 15% methanol over 11 minutes, followed by a more rapid linear increase to 45% methanol over 5 minutes. Solvent was removed from the fractions using rotary evaporation, and resulting samples were dissolved in methanol-d_4 and analyzed by ^{1}H NMR spectroscopy. The fraction containing **10** (eluting at 8-9% methanol) was purified further using reverse-phase HPLC. The sample was dissolved in 500 µL of methanol, and 20 µL of this sample was injected onto the semi-preprative column running a water/methanol gradient (3 mL/min) starting with 5% methanol for 5 minutes, followed by a linear increase to 100% methanol over 50 minutes, and holding at 100% methanol for 15 minutes. At the end of the separation, the column was equilibrated at 5% methanol for 15 minutes, followed by reinjection of sample. This process was repeated until the entire sample was chromatographed. After evaporation of solvent, fractions containing **10** (eluting at ~90% methanol) combined and characterized *via* ^{1}H, dqfCOSY, ROESY, ^{1}H,^{13}C-HSQCAD, and ^{1}H,^{13}C-gHMBCAD (dqfCOSY and ROESY spectra were acquired in both methanol-d_4 and DMSO-d_6).

10. Heterologous lnaA- and lnbA-Expression: For heterologous production of *N*-terminally tagged hexahistidine fusion proteins, *E. coli* soluBL (Genlantis) was

transformed with plasmids pSA14 and pSA15, carrying the NPRS genes *lnaA* and *lnbA*, respectively. *E. coli* x pSA14 was grown as an overnight seed culture in LB-medium (10 mL), which was used to inoculate 500 mL of autoinduction medium as production culture. Ampicillin (100 µg/mL final) was added to all cultures. The production culture was incubated at 180 rpm and 37 °C until an OD_{600} of 0.6 - 0.8 was reached. The temperature was then lowered to 22 °C, shaking at 160 rpm, for further 12 h. *E. coli* x pSA15 was grown as described above, but in LB-medium as production culture. Expression was induced with 1 mM IPTG. **(A) Protein purification**. Cells were harvested by centrifugation (4 °C, 4.000 x g, 25 min). The cell paste was resuspended in 10 mL wash buffer (50 mM sodium phosphate / 300 mM sodium chloride, pH 8.0), amended with 20 mM imidazole, and the cells disrupted by a Branson sonifier (4 x 40 s pulses, pause between pulses: 60 s). Debris was removed by centrifugation (4 °C, 6.000 x g, 25 min), and Ni-NTA slurry was added to the cleared supernatant and stirred on ice for 1 h. Purification was performed according to the manufacturer´s instruction (QIAGEN expressionist), except that the wash steps involved a step gradient from 50-10 mM imidazole and that elution was done with 500 mM imidazole. The first elution fraction (ca. 1 mL) was discarded, the other fractions collected. Removal of imidazole and desalting was accomplished using a PD-10 column (GE healthcare) and assay buffer (below).

(B) Biochemical characterization of *Aspergillus flavus* LnaA and LnbA. The ATP-[^{32}P] pyrophosphate exchange assay was carried out as described[9,10] with the following modifications: i) assay buffer was sodium phosphate 250 mM; $MgCl_2$ 12.5 mM; ATP 12.5 mM; pH 6,5; ii) the wash steps included 200 mL wash solution, 200 mL water, and 35 mL ethanol; iii) the substrates were provided as 4 mM solution. Scintillation was read out on a Beckman Coulter LS6500 counter. Chemicals and media components were purchased from Becton-Dickinson, Fisher, Fluka, Roth, and Sigma-Aldrich, except sodium [^{32}P]pyrophosphate (activity: 62.9 Ci/mmol) which was from PerkinElmer.

Substrate specificity was assessed with five pools of L-amino acids and a sixth pool including 2-oxo acids: Pool 1: Ala, Gly, Leu, Ile, Val; Pool 2: Cys, Met, Pro, Ser, Thr; Pool 3: His, Phe, Tyr, Trp; Pool 4: Asp, Asn, Glu, Gln; Pool 5: Arg, Lys, Orn; Pool 6: phenylpyruvic acid, 4-hydroxyphenylpyruvic acid, indolyl-3-pyruvic acid, pyruvic acid, α-keto glutaric acid. Water served as negative control throughout. For assays with pure amino acids, L-Ala was chosen as additional negative control.

11. HPLC-MS-Based Profiling of 1-8 in the ΔlnaA KD::lnbA, ΔlnaA, Wild-Type, OE::lnaA KD::lnaB, and OE::lnaA A. flavus Strains: For comparing the relative abundances of **1-8** between different strains and conditions, cultures to be compared were grown on the same media, at the same time, and were processed and extracted under the same conditions. For measuring compounds **1** and **2** and studies comparing *OE::lnaA* and *OE::lnaA, KD::lnaB*, three biological replicates comprised of a pooling of 10 GMM cultures each were used. Unless otherwise specified, HPLC-MS strain profiling was performed on extracts derived from cultures grown on GMM under conditions that support expression of the *lna* and *lnb* gene clusters. **(A) HPLC chromatography**. Samples were dissolved in 0.2-1.0 mL of methanol. If a precipitate formed upon addition of methanol, the sample was filtered through a methanol-washed cotton filter. 2.0 µL of crude extract sample was injected into an Agilent 1100 series HPLC equipped with a Zorbax Eclipse XDB-C8 column, using a flow rate of 2.0 mL/min, and a water (0.1% acetic acid) / methanol gradient starting with 1% methanol for 10 minutes followed by a linear increase over 20 minutes to 100% methanol, and ending with 10 minutes at 100.0% methanol. Synthetic standards of **1** and **2** were used to determine which chromatographic peaks represent **1** or **2**, and enriched samples of compounds **1**-**8** served as retention time standards for the remaining analytes.

(B) Electrospray positive ionization (ESI^+) mass spectrometry. The eluent of the HPLC was split into a Quattro II triple quadrupole mass spectrometer operated in

either single ion monitoring (SIM) or centroid (*m*/z range 2.0–800.0) acquisition modes. See **Figure C.11** for elution profile of **1**-**8**.

(C) Estimating production of 1 and 2 in the Δ*lnaA*, Wild-type and *OE::lnaA* strains. Synthetic standards of **1** and **2** were dissolved in 0.6 mL of methanol-d_4 (99.8% D), and quantitative ^{1}H spectra were acquired. After baseline correction, the doublet of doublets belonging to one of the benzylic hydrogens of **1** or **2** was integrated relative to the C<u>H</u>D_2OD solvent signal. Using the C<u>H</u>D_2OD signal as an internal standard (representing 0.2% of the 0.6 mL methanol-d_4 sample) the amount **1** and **2** in the NMR sample was estimated (**1**, 0.2 mg; **2** 0.6 mg). Solvent was removed from the standards using rotary evaporation and 1.0 mL of methanol was added. These standards were diluted up to 10^6-fold and analyzed by HPLC/ESI$^+$-SIMMS, indicating roughly linear spectrometer response over the entire concentration range. Extracts of Δ*lnaA*, wild-type and *OE::lnaA* strains (derived from 20 GMM plates each) were dissolved in 1.0 mL methanol and analyzed by HPLC/ESI$^+$-SIMMS. This analysis showed: Δ*lnaA* **1**, 1 pmol/plate; Δ*lnaA* **2**, 9 pmol/plate; wild-type **1**, 74 pmol/plate; wild-type **2**, 97 pmol/plate; *OE::lnaA* **1**, 560 pmol/plate; *OE::lnaA* **2**, 600 pmol/plate.

(D) Data workup. Ion chromatographic data were processed in MassLynx. Peak areas were measured in the HPLC/ESI$^+$-(SIM)MS ion chromatograms. *(1) Measuring the compound **1**-to-compound **2** value:* For measuring the compound **1**-to-compound **2** ratio for the Δ*lnaA*, wild-type and *OE::lnaA* strains, SIMMS ion chromatogram peak areas (ion monitored, 299.2 *m*/z) were measured. Given that the Δ*lnaA* strain produces ~ 10 fold less **1** and **2** than wild-type (as determined by ion chromatogram peak areas), the wild-type extracts were diluted by a factor of 10 and rerun to determine if the decrease in the wild-type **1** and **2** S/N would significantly affect the measured compound **1**-to-compound **2** ratio, which was not the case. This additional validation step was also used to re-analyze the compound **1**-to-compound **2** ratio in the *OE::lnaA* extracts. Results from the Δ*lnaA*, wild-type, and *OE::lnaA* compound **1**-to-compound **2**

study were reproduced using a second, separately grown and extracted culture set. *(2) Profiling the OE::lnaA, KD::lnaB and OE::lnaA strains:* To measure changes in *lna*-dependent metabolite profile as a result of *lnaB* knock down, compounds **1**-**8** were profiled for *OE::lnaA, KD::lnaB* and *OE::lnaA* extracts using HPLC/ESI$^+$-SIMMS. This profiling method was also used to profile the presence or absence of compounds **1**-**8** in cultures of Δ*lnaA*, wild-type, and *OE::lnaA* grown on rich media and GMM.

12. Wild-Type A. flavus 1 and 2 Feeding and lna Cluster Expression Experiment: Aqueous solutions (0.1 % acetic acid) of synthetic **1** and **2** (0.4 mM) were sterile-filtered using a 0.2 µm Thermo Scientific Nalgene syringe filter. A spore suspension of wild-type conidia was prepared as described in the main text methods, diluted, and used to inoculate in 20 mL liquid GMM cultures. 10 plates each were treated with either 300 µL of **1** solution, **2** solution, or vehicle control. The cultures were incubated at 29 °C in the dark as described above. Mycelia were collected after 6 days and portions were used to extract total RNA. RNA expression was assessed as described in Appendix C, section 1(E).

13. HRMS Analysis of Compounds 1, 2, 4-7, 10, 19, 20†, 22, and 23: Enriched samples of **4**-**6** were dissolved in 0.2-0.5 mL of methanol and injected into the UPLC/ESI$^+$-HRMS system described in the main text methods. Various chromatographic solvent gradients were used; all based on gradients of 5-100% methanol in water containing 0.1% formic acid. Mass spectra were acquired in ESI$^+$ and, for **4** and **6**, ESI$^-$ modes. High resolution mass spectra were acquired for compounds **1**, **2**, **7**, **10**, **19**, **20**, **22**, and **23** by dissolving the purified samples (~0.5 mg) in 1.0 mL of methanol and diluting an aliquot of the samples by a factor of 100 with methanol containing 0.1% formic acid. The resulting solutions were directly infused into an LTQ Orbitrap Velos mass spectrometer operating in ESI$^+$ ionization mode. **7** was

additionally incubated in methanol-d_4 resulting in the exchange of four H atoms for four D atoms (see table below).

Compound	**ESI$^{+/-}$ HRMS Observed *m*/z**	**Ion**	**Calculated Ion Formula**	**Calculated *m*/z**
1	299.1751	$[M + H]^+$	$C_{18}H_{23}N_2O_2^+$	299.1754
2	299.1754	$[M + H]^+$	$C_{18}H_{23}N_2O_2^+$	299.1754
4	371.0703	$[M - H]^-$	$C_{18}H_{15}N_2O_5S^-$	371.0707
5*	309.1236	$[M - O + H]^+$	$C_{18}H_{17}N_2O_3^+$	309.1234
5*	291.1133	$[M - O - OH-]^+$	$C_{18}H_{15}N_2O_2^+$	291.1128
6	373.0865	$[M + H]^+$	$C_{18}H_{17}N_2O_5S^+$	373.0853
6	371.0703	$[M - H]^-$	$C_{18}H_{15}N_2O_5S^-$	371.0707
7	316.1546	$[M + H]^+$	$C_{18}H_{22}NO_4^+$	316.1543
7-d_1 (after incubation in methanol-d_4)	317.1605	$[M + H]^+$	$C_{18}H_{21}DNO_4^+$	317.1606
7-d_4(4H atoms exchanged, D)	320.1788	$[M + H]^+$	$C_{18}H_{18}D_4NO_4^+$	320.1794
10	474.2628	$[M + Na]^+$	$C_{28}H_{37}NNaO_4^+$	474.2615
19	543.2191	$[M + H]^+$	$C_{24}H_{39}N_4O_6Zn^+$	543.2156
20†	511.2	$[M + H]^+$	n/a	n/a
20 (**21** frag.)	225.1	$[M + H]^+$	n/a	n/a
22	774.3618	$[M + H]^+$	$C_{36}H_{58}FeN_6O_9^+$	774.3609
23	726.3762	$[M + H]^+$	$C_{36}H_{58}FeN_6O_6^+$	726.3762

* See reference[11] for ionization routes for aromatic *N*-oxides.

† Unit-resolution mass spectrum acquired using triple quadrupole MS.

14. Sclerotia Analysis: Because the *lna* and *lnb* clusters were found to be regulated by LaeA and highly expressed during sclerotia-forming conditions, we compared the *lnaA* and *lnbA* mutant strains with wild-type for sclerotial production. Sclerotia formation exhibits a quorum-like density-dependent growth pattern in *A. flavus* whereby sclerotia are abundantly produced at low population densities but suppressed at high population densities. For sclerotia analysis, plates were sprayed with 70% ethanol to remove conidia, and sclerotia were scraped off using a small spatula. The sclerotia were placed into pre-weighed microcentrifuge tubes, flash-frozen in liquid nitrogen, lyophilized, and weighed again, and the weight of the sclerotia was calculated by subtracting the initial weight of the empty tube from the final weight. The data were analyzed using Prism 5 software to carry out a one-way ANOVA with a Tukey post-test and a p-value of <0.05. 60-mm diameter plates containing 8 mL GMM + 2% sorbitol and 15 g/L agar were overlaid with 3 mL GMM + 2% sorbitol and 7.5 g/L agar containing 10^3 or 10^7 spores. Five replicates per set were placed in the dark at 29°C for seven days. Sclerotial formation was assessed at 10^3 spores/plate (low density) and 10^7 (high density). All tested mutant strains were aberrant in sclerotial development, including total weight, size and pattern of production (**Figure C.1**). With the exception of the Δ*lnaA, KD::lnbA* strain, mutant strains showed abundant sclerotia production at low population densities, similar to wild-type. However, all strains produced significantly more sclerotia than wild-type at high population densities. Notably, the Δ*lnaA, KD::lnbA* strain TSA22.1 exhibited a phenotype in which sclerotia production was suppressed at low rather than high population densities, opposite of what is observed in wild-type (**Figure C.1**).

15. Compound 1-to-Compound 2 Ratio OE::lnaA, Wild-Type,* Δ*lnaA: As deletion of *lnaA* decreased the compound **1**-to-compound **2** ratio, the compound **1**-to-compound **2** ratio was measured for *OE::lnaA*, expecting a higher **1** to **2** ratio relative to wild-type. However, it was found that the *OE::lnaA* compound **1**-to-compound **2** ratio was not

increased relative to wild-type, but instead placed between that of wild-type and Δ*lnaA* (**Figure C.6**). Given that the *OE::lnaA* construct was shown to increase expression of both *lnaA* and *lnaB*, and both proteins contain a putative epimerase domain, it is possible that the compound **1**-to-compound **2** ratio observed for *OE::lnaA* results in part from increased LnaB expression.

16. Statistical Analysis: Statistical differences were analyzed using the JMP software package (version 9.0.2, SAS Institute, Inc, Cary, NC). Multiple comparisons of all strains were calculated for diameter growth, conidia production. Mean values with different letters or asterisk are significant in the group

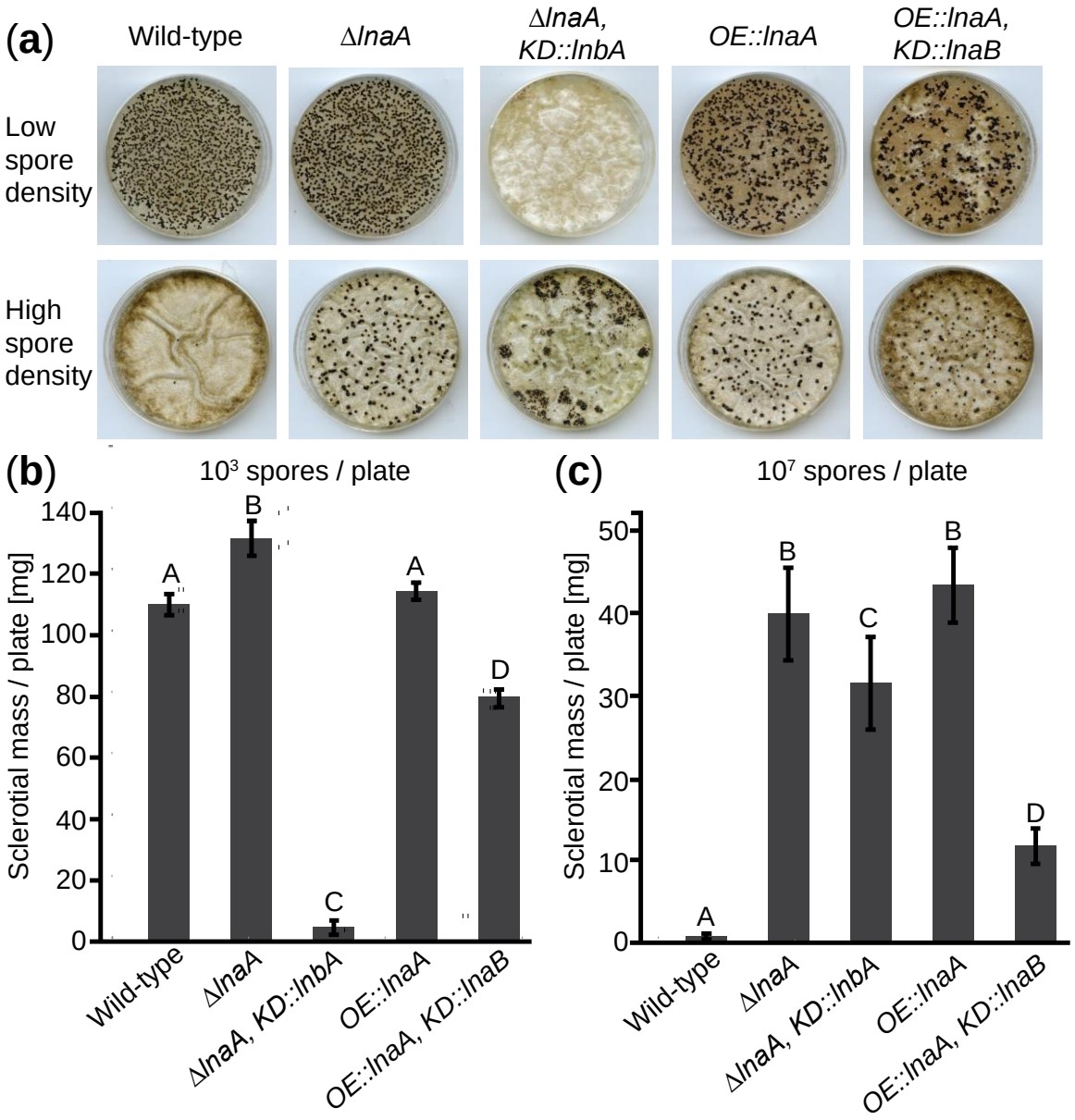


Figure C.1: Sclerotia production by *lna* and *lnb* cluster mutants. (**a**) Images of plates after washing off conidia. The sclerotia are the dark structures left on the plate after conidia are washed off. (**b**) Quantification of sclerotia formed on plates inoculated with 10^3 and (**c**) 10^7 spores. Different letters represent significantly different groups with a p-value < 0.05, according to ANOVA analysis followed by a Tukey post-test, and error bars signify one standard deviation (n = 5).

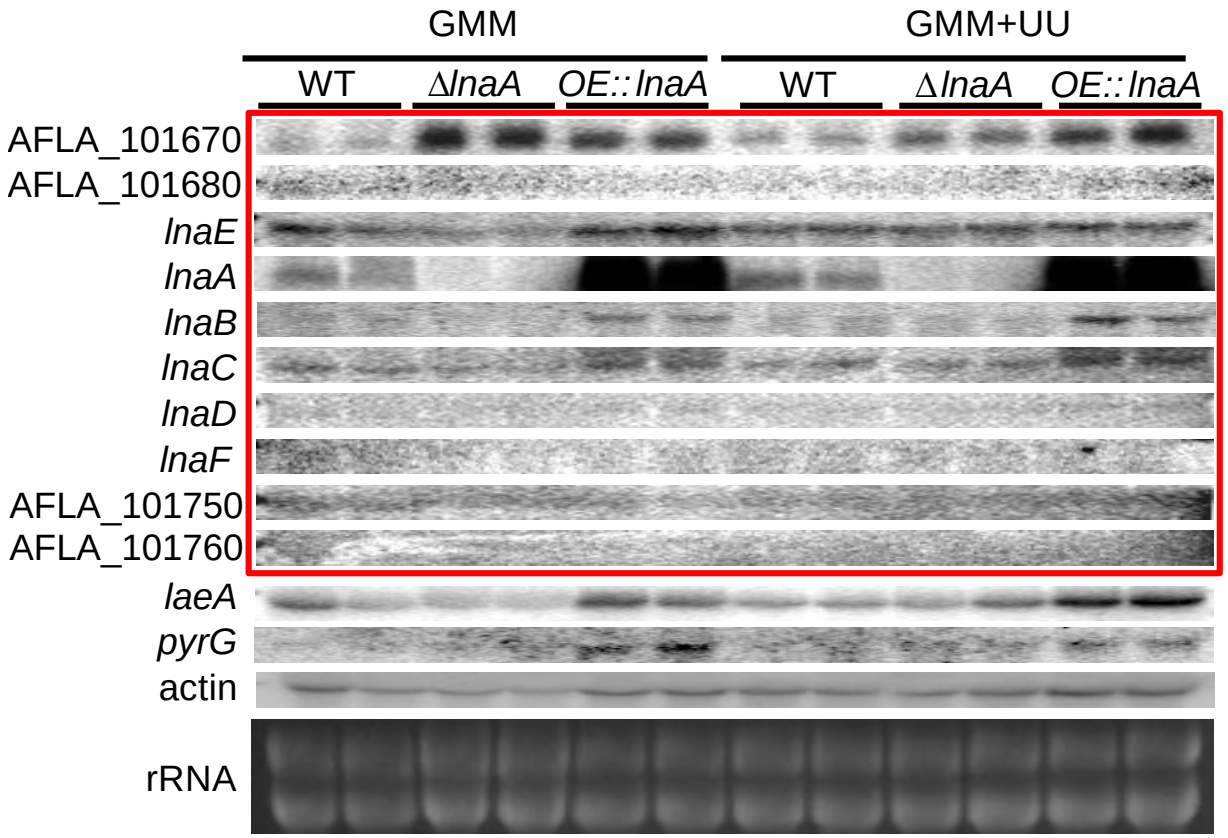


Figure C.2: Northern expression analysis for the wild-type (WT), Δ*lnaA*, and *OE::lnaA* strains. GMM (glucose minimal media) GMM+UU (GGM with uracil and uridine) growth conditions were compared for expression influence on the *lna* gene cluster and surrounding genes. Genes listed on the left correspond to: AFLA_101670, AFLA_101680, *lnaE* (P450, AFLA_101690), *lnaA* (NRPS, AFLA_101700), *lnaB* (*nmrA*-like gene, AFLA_101710), *lnaC* (dehydrogenase, AFLA_101720), *lnaD* (P450, AFLA_101730), *lnaF* (transporter, AFLA_101740), AFLA_101750, and AFLA_101760. Actin and rRNA serve as loading controls.

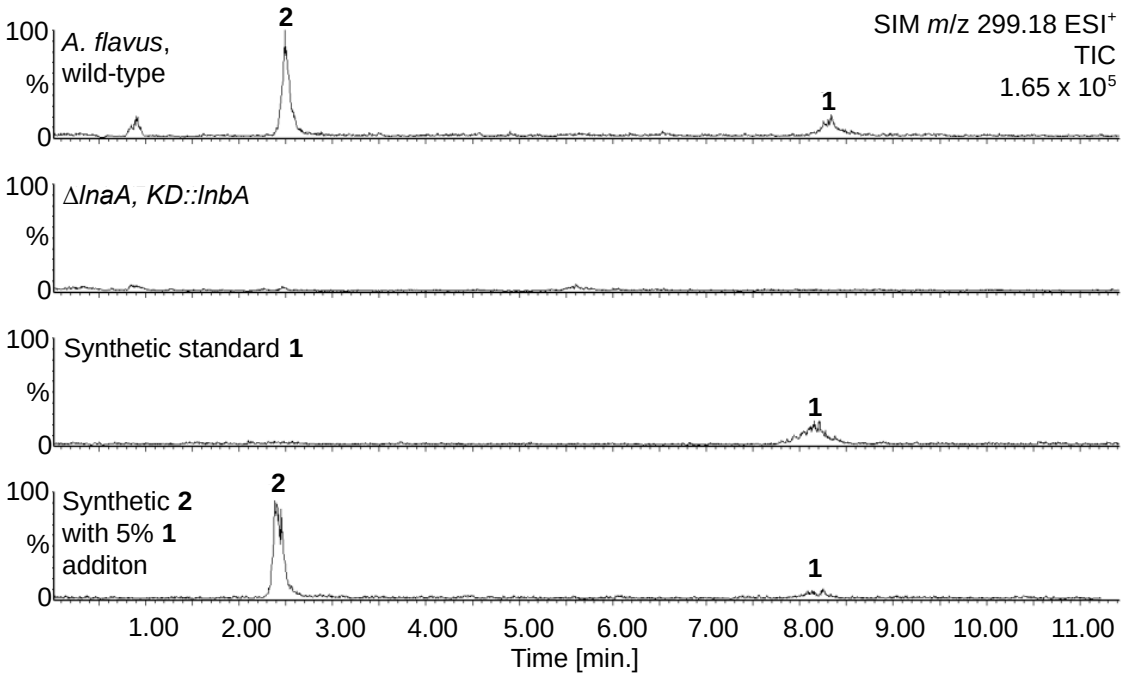


Figure C.3: Analysis of the Δ*lnaA*, *KD::lnbA* strain for the presence of absence of compounds 1 and 2. SIM ion chromatograms of *m*/z 299.18 for extracts corresponding to the wild-type and Δ*lna*, *KD::lnbA* (TSA22.1) strains as well as synthetic standards of **1** and **2**. The Δ*lna*, *KD::lnbA* strain showed no evidence for the presence of **1** nor **2**. Analysis of the second Δ*lna*, *KD::lnbA* strain (TSA22.6) gave the same result.

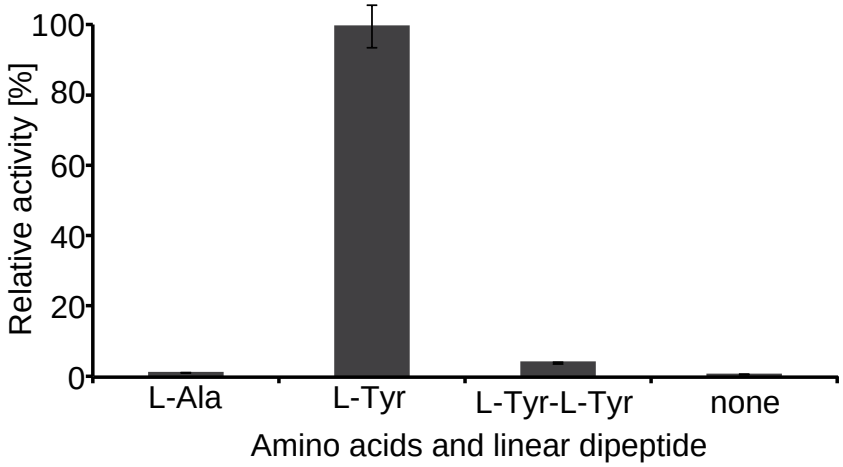


Figure C.4: LnaA does not show substrate specificity for linear L-Tyr-L-Tyr. Recombinant LnaA was tested for its ability to adenylate L-Ala, L-Tyr, and L-Tyr-L-Tyr. The bar diagram shows the relative activity values, as determined *via* the amino acid-dependent ATP [^{32}P]-pyrophosphate exchange assay, for each substrate acid tested, referenced to the maximum turnover (L-Tyr). Error bars signify +/- one standard deviation (n = 3).

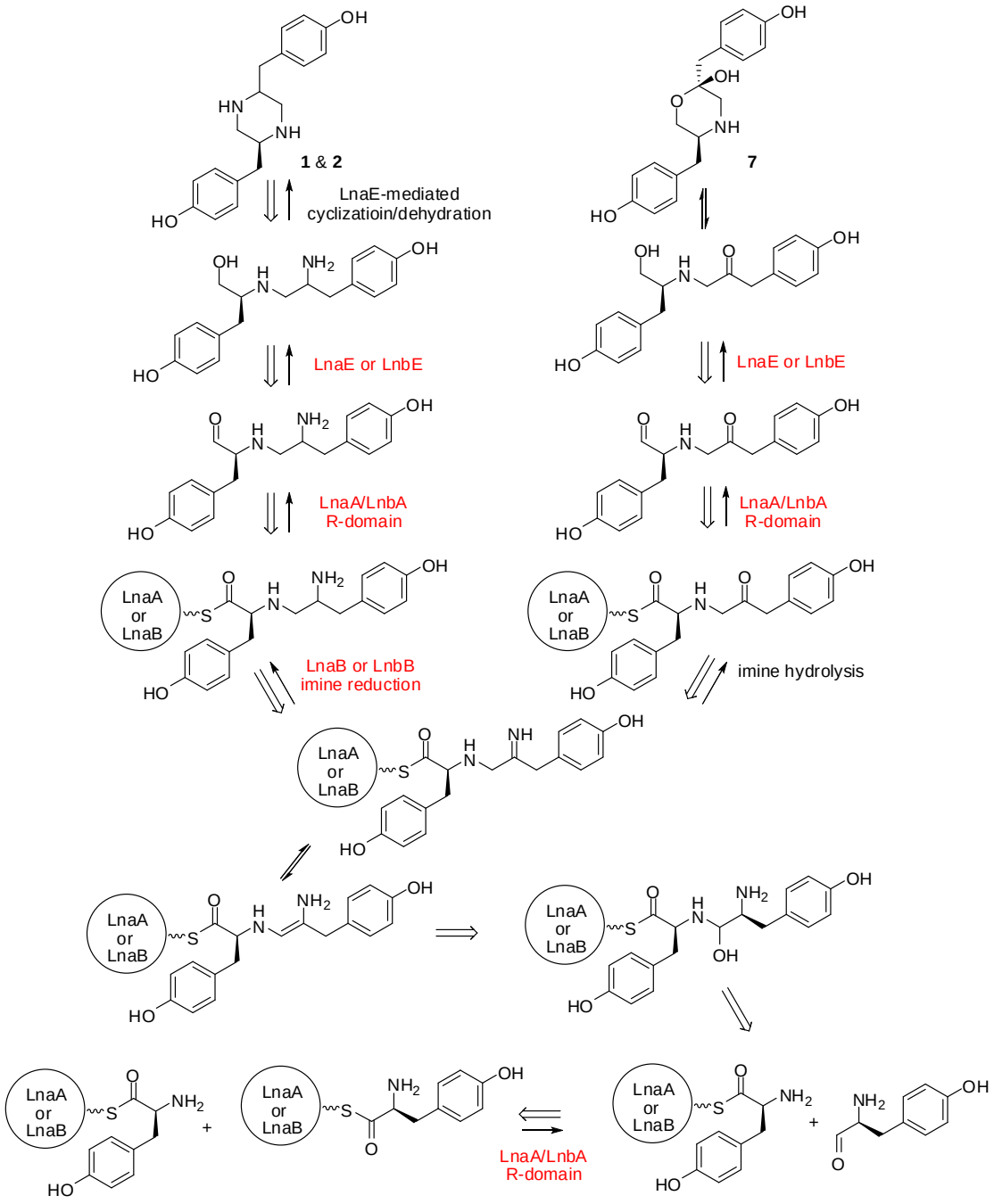


Figure C.5: Retro biosynthetic scheme for compounds 1, 2, and 7. Forward reaction steps with red lettering represent reductive steps in the biosynthesis. Note that the four reductive steps are required for the production of **1** or **2**, whereas only 3 reductive steps are required for the generation of **7.** This model accommodates the observation that knocking down the expression of *lnaB*, encoding a putative NmrA-like reductase, results in a shift from **1** and **2** production towards greater production of **7**.

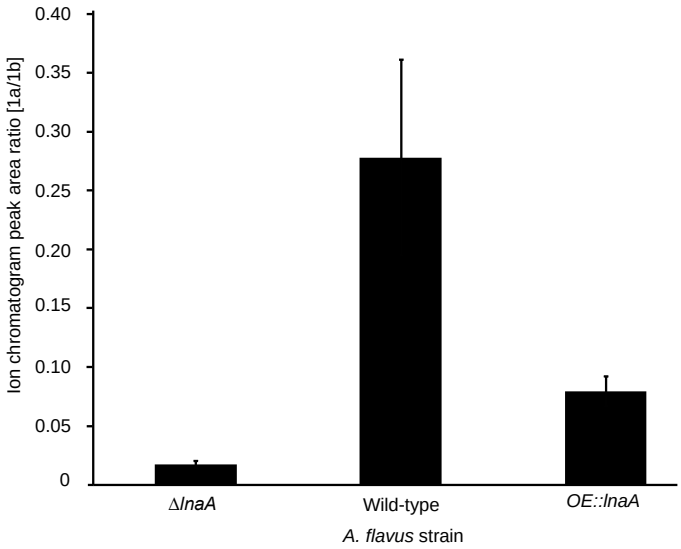


Figure C.6: Ion chromatogram peak area ratios compound 1-to-compound 2 for Δ*lnaA*, wild-type, and *OE::lnaA* (TSA7.15). Peak areas were determined from integration of HPLC/ESI$^+$-SIMMS ion chromatograms (*m*/z 299.18) corresponding to three replicates for each strain. Error bars signify +/- one standard deviation (n = 3). As demonstrated previously the wild-type compound **1**-to-compound **2** ratio is greater than that of Δ*lnaA*, whereas the *OE::lnaA* mutation lowers the compound **1**-to-compound **2** ratio to a value between that of wild-type and Δ*lnaA*.

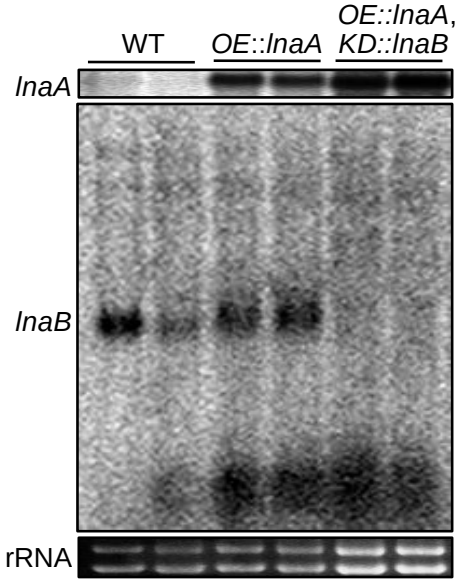


Figure C.7: Northern analysis of wild-type (WT), *OE::lnaA*, and *OE::lnaA, KD::lnaB.* *A. flavus* WT, *OE::lnaA*, and *OE::lnaA, KD::lnaB* strains were examined by Northern analysis with *lnaA* and *lnaB* probes. The loss of *lnaB* expression in the latter strain confirms silencing of *lnaB.*

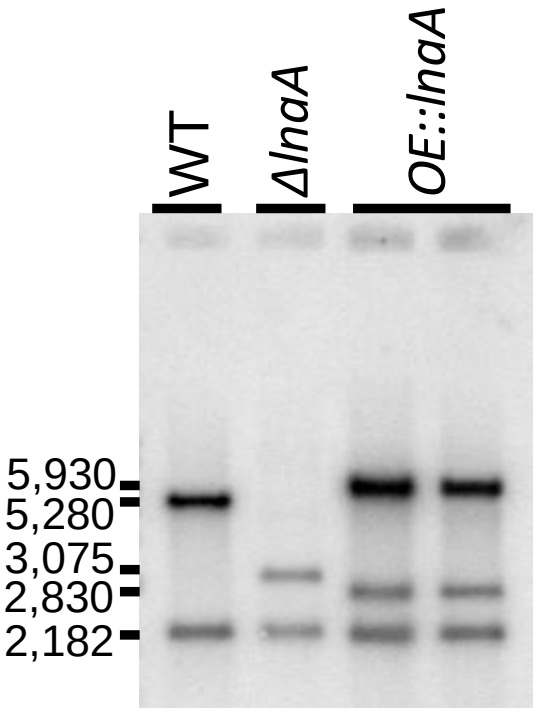


Figure C.8: Southern analysis of wild-type (WT) and *lna* gene cluster mutants Δ*lnaA* and *OE::lnaA*. Each sample was digested with NdeI and probed with the ORF of *lnaA*. WT: 5280bp, 2182bp, *ΔlnaA*: 3075bp, 2182bp, *OElnaA*: 5980bp, 2830bp, 2182bp. Bands are consistent with predicted restriction sites.

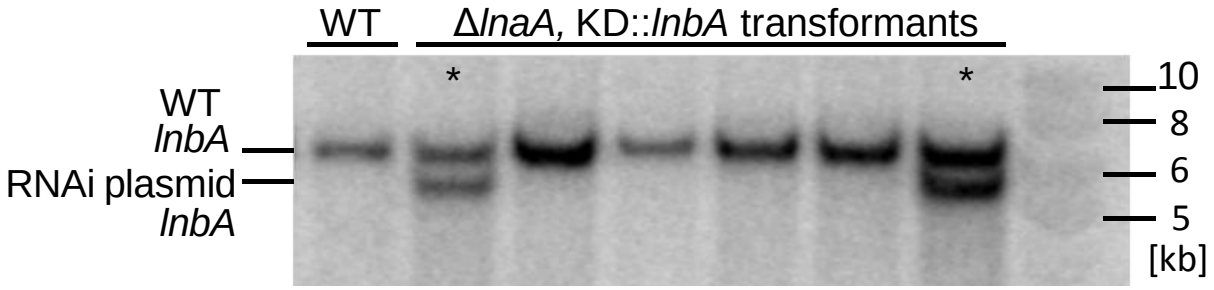


Figure C.9: TSA22.1/22.6 strains confirmation by Southern analysis. Lane one, wild-type (WT) strain NRRL2257, in lanes 2-7 "*" indicates TSA22.1 and TSA22.6, which showed extra copy of *lnbA* from successful incorporation of the RNAi plasmid, pSA7.3.

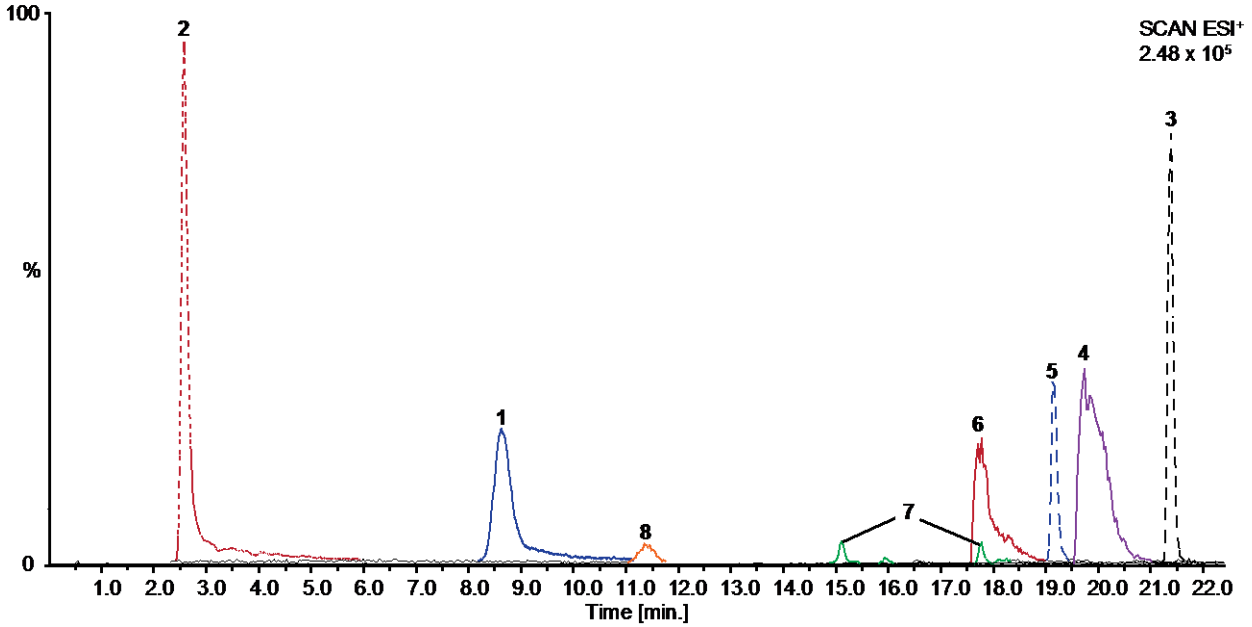


Figure C.10: Overlay of ion chromatograms corresponding to compounds 1-8. Traces were derived from HPLC/ESI$^+$-MS chromatograms acquired for enriched samples of **3**-**8** (traces for **1** and **2** are derived from single ion monitoring traces acquired for a crude wild-type extract). HPLC/ESI$^+$-MS chromatographic and mass spectrometry acquisition parameters are outlined in Supporting Information section 10(A). Peak intensities are scaled relative to the peak in the ion chromatogram with the largest intensity, **2**. Compound **2**, red dotted trace; **1**, solid blue trace; **8**, solid orange trace; **7** and isomers, solid green trace; **6** solid red trace; **5** blue dotted trace; **4** solid violate trace; and **3**, black dotted trace.

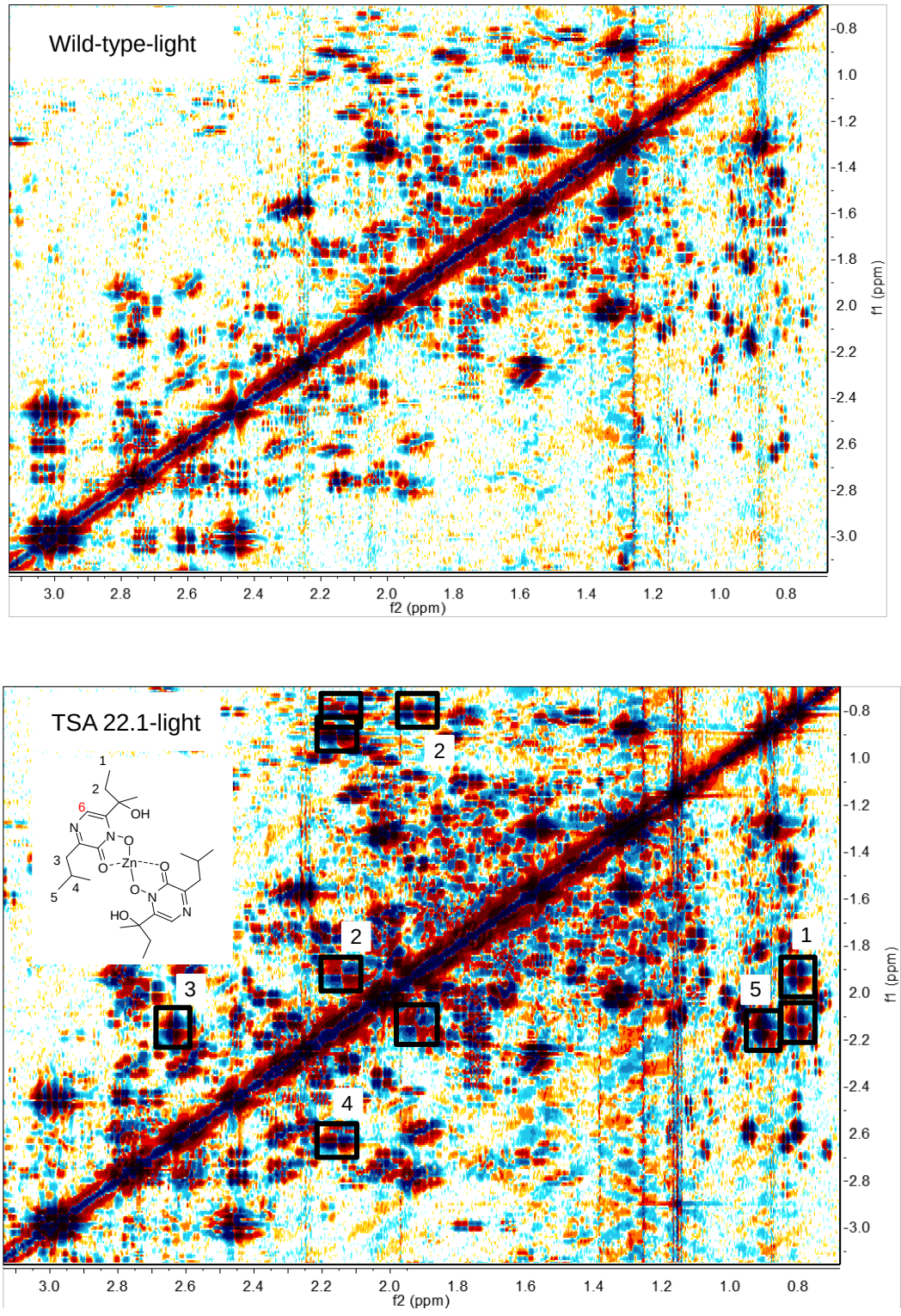


Figure C.11 Comparison of dqfCOSY spectra for wild-type-light and TSA 22.1-light. TSA 22.1-light shows dramatic increase in the production of compound **19**.

Table C.1: Gene symbols and names used in this study as well as annotated functions (www.pubmed.org).

Gene symbol	Gene name	Annotated function
AFLA_101690	*lnaE*	Alcohol dehydrogenase/NAD:flavin oxidoreductase
AFLA_101700	*lnaA*	NRPS
AFLA_101710	*lnaB*	*nmrA*-like gene (putative reductase/epimerase)
AFLA_101720	*lnaC*	Cytochrome P450
AFLA_101730	*lnaD*	Cytochrome P450
AFLA_101740	*lnaF*	Transporter
AFLA_121470	Not named	*nmrA*-like gene (putative reductase/epimerase)
AFLA_121480	Not named	dioxygenase
AFLA_121490	*lnbB*	*nmrA*-like gene (putative reductase/epimerase)
AFLA_121500	*lnbC*	Cytochrome P450
AFLA_121510	n/a	methyltransferase
AFLA_121520	*lnbA*	NRPS
AFLA_121530	*lnbE*	Alcohol dehydrogenase/NAD:flavin oxidoreductase
AFLA_121540	*lnbF*	Transporter

Table C.2: List of primers used in this study.

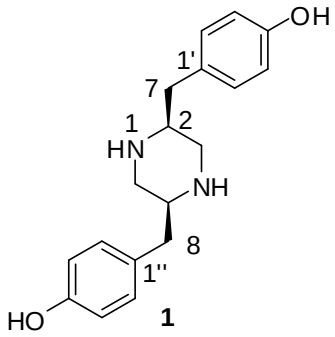
Name	**Sequence 5'-3'**
DM101700-5' For	GGT CGC GTC GTC AGC TTT AAT G
DM101700-5' Rev	CGA AGA GGG TGA AGA GCA TTG TTT GAG GCA CTA ACC TCC ATC CGC GAC AC
A. fumigatus pyrG For	TGCCTCAAACAATGCTCTTC
A. fumigatus pyrG Rev	CAAGGTATCGTCGGGAGGT
DM101700-3' For	GTG ACG ACA ATA CCT CCC GAC GAT ACC TGG GAC AAG CGG AAG AAG GGA GT
DM101700-3' Rev	CAT GCG TAT CAA AGGG CCT CCC A
DM101700 nested For	GAC TTG CGC GAT ACT CAT CGT C
DM101700 nested Rev	GCT GAC ACC GTG AAC GAA GTG A
OE101700-5' For	CAA AAC TGC GGA TGT CGC TGA C
OE101700-5' Rev	CGA AGA GGG TGA AGA GCA TTG TTT GAG GCA CGA GGA TTT CGT TTT CCA GG
A. nidulans gpdA(p)For	AAG GCT TG GGC CGC TGC GTT GGT T
gpdA(p)::pyrG For	GTG ACG ACA ATA CCT CCC GAC ACC TGG CAT CCG GAT GTC GAA GGC TTG
A. nidulans gpdA(p) Rev	CAT GGT GAT GTC TGC TCA AG
OE101700-ORF For	AGC TAC CCC GCT TGA GCA GAC ATC ACC ATG CCT TCT CAA GTT TTG ACT CA
OE101700-ORF Rev	GCC TCG AGC TCA ATT CGA AAC C
Seq101700 For1	GTT GAC AAG GTC GTT GCG TCA G
NSeq101700 For2	CAT GGA TCC AAC CTT GCC TGA C
Seq101700 For3	TGA TGA AGC TGG TCG ACC AGT C
Seq101700 Rev1	CGA GCT CTT GGA AAA TGA CGC G
Seq101700 Rev2	GAT TGA TGT ATC CGG GCG AGA C
NSeq101700 Rev3	CGA GCC TGA TGA GGC TGC TAT T
DM101700 For	GGC CTT GCC CAA CAA TTC TTC GAT
DM101700 Rev	CAG CAA GCA AGA AAG AGC CCA CAA
IRT101710 5F For	CTA CCC CGC TTG AGC AGA CAT CAC CA
IRT101710 5F Rev	CTC TCT CCT GCA CAT AGC CCT CGG CG
IRT101710 3F For HindIII	GAA GAA CAA GCT TGA GCA GTA CGC AA
IRT101710 3F Rev NotI	CTA GTT TGC GGC CGC AGC AAA CCA GT
pTMH44 seq1For	CTA CAT CCA TAC TCC ATC CTT C
pTMH44 seq1Rev	GTG GCC GAG AAT GTT TCC ATC C

pTMH44 seq2For	CAA GTT CGA AGG TGA CAC CCT G
pTMH44 seq2Rev	CCA TTT GTC TCA ACT CCG GAG C
IRT121520 5F For	CTA CCC CGC TTG AGC AGA CAT CAC CA
IRT121520 5F Rev	CTC TCT CCT GCA CAT AGC CCT CGG CG
IRT121520 3F For BamHI	GAA CGA GAG GAT CCA AAG CTC CCC TT
IRT121520 3F Rev NotI	GAT TGA CGC GGC CGC TCG ATC GAT CT
IRTlaeA 5F For	CTA CCC CGC TTG AGC AGA CAT CAC CA
IRTlaeA 5F Rev	CTC TCC TGC ACA TAG CCC TCG GCG C
IRTlaeA 3F For BamHI	GAA CGA GAG GAT CCG TCT TGA TGC CA
IRTlaeA 3F Rev NotI	GAT TGA CGC GGC CGC CCG ATT TCT AG
1)NlaeA For	CCT TGT ATG ATG TAT GTA TGA TGA GC
NlaeA Rev	GAC AGC GAA AGT GAA GAG GAC ATC
Nactin For	GAAGCGGTCTGAATCTCCTG
Nactin Rev	ACAGTCCAAGCGTGGTATCC
N101670 For	CTA TCC GAC ATT GAA AAG GTC GC
N101670 Rev	CCG AGA TAC AAT ACA CTT TGC TGA
N101680 For	GAT TTT CCT TGA GGC CAA TAG TTC
N101680 Rev	CAC GTT ACC ATT TTA TCT ACC TGC TG
N101690 (lnaC) For	GGA TCA GGA CAC GGA GAA CAA G
N101690 (lnaC) Rev	CGT AAC CGC TAA GAT GGT ATG CG
N101710 (lnaB) For	CTT GGA CGA CAT CGT GGG AAT G
N101710 (lnaB) Rev	CGT ATC CGT ATA GGC CTA CCC T
N101720 (lnaD) For	CGT AGT TGT CGA ATG CTG GAG C
N101720 (lnaD) Rev	GTC TCA GCG CCA ATT TCT GTC G
N101730 (lnaE) For	CAA GGA GCG AGC GTA TCC TTT C
N101730 (lnaE) Rev	GTC TCC TTC ACC ACA CTG ATG G
N101740 (lnaF) For	CAA TGG GAG AAG CTA GGG CCG ATG
N101740 (lnaF) Rev	GGA GAC AAA TAG TGG TCC ACG G
N101750 For	GGC TCA AGC TCC GAC CGG GAC
N101750 Rev	CCC GAA CTA ACA GGG GGA ATG A
N101760 For	TCC AAC ATT ACA GCC CGC AAT AC
N101760 Rev	CTA GCT CGA TAA GCT GGC ACA G

Table C.3: *A. flavus* strains, Abbreviations, and Genotypes.

Name	Abbreviation	Genotype	References
NRRL 3357	WT	wild-type	[12]
NRRL 3357.5	*pyrG*-WT	*pyrG*	[12]
TSA 5.21	Δ*lnaA*	*pyrG*-, *ΔAFLA_101700::A. fumigatus pyrG*	This study
TSA 7.15	*OE::lnaA*	*pyrG*-, *gpdA*(p)*::AFLA_101700::A. fumigatus pyrG*	This study
TSA 7.51	*OE::lnaA*	*pyrG*-, *gpdA*(p)*::AFLA_101700::A. fumigatus pyrG*	This study
TSA 22.1	Δ*lnaA*, *KD::lnbA*	*pyrG*-, *A. fumigatus pyrG::AFLA_101700, IRT AFLA_121520, phleomycin R*	This study
TSA 22.6	Δ*lnaA*, *KD::lnbA*	*pyrG*-, *A. fumigatus pyrG::AFLA_101700, IRT AFLA_121520, phleomycin R*	This study
TSA 16.7.31	Δ*lnaA*, *KD::lnaB*	*pyrG-, A. fumigatus pyrG*::*A. nidulans gpdA*(p)::*lnaA*, IRTl*naB*, phleomycin	This study

Table C.4: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **1** in methanol-d_4. Chemical shifts were referenced to δ(C<u>H</u>D$_2$OD) = 3.31 ppm and δ(^{13}<u>C</u>HD$_2$OD) = 49.0 ppm.^{13}C chemical shifts were determined via HMBC spectroscopy. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.

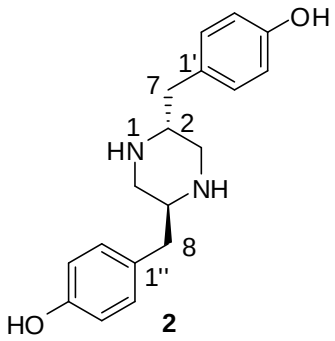


1

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC
1		1-NH		
2	52.6	2-H	2.87 ($J_{2\text{-H, 3-Ha}}$ = 4, $J_{2\text{-H, 3-Hb}}$ = 7, $J_{2\text{-H, 7-Ha}}$ = 8, $J_{2\text{-H, 7-Hb}}$ = 7)	3, 7, 1'
3	43.4	3-H$_a$	2.72 ($J_{3\text{-Ha, 3-Hb}}$ = 12)	2, 7
		3-H$_b$	2.75	2, 7
4		4-NH		
5	52.6	5-H	2.87 ($J_{5\text{-H, 6-Ha}}$ = 4, $J_{5\text{-H, 6-Hb}}$ = 7, $J_{5\text{-H, 8-Ha}}$ = 8, $J_{5\text{-H, 8-Hb}}$ = 7)	6, 8, 1''
6	43.4	6-H$_a$	2.72 ($J_{6\text{-Ha, 6-Hb}}$ = 12)	5, 8
		6-H$_b$	2.75	5, 8
7	34.4	7-H$_a$	2.73 ($J_{7\text{-Ha, 7-Hb}}$ = 13, $J_{7\text{-Ha, 2'-H}}$ < 1, $J_{7\text{-Ha, 6'-H}}$ < 1)	3, 1', 2', 6'
		7-H$_b$	2.77 ($J_{7\text{-Hb, 2'-H}}$ < 1, $J_{7\text{-Hb, 6'-H}}$ < 1)	3, 1', 2', 6'
8	34.4	8-H$_a$	2.73 ($J_{8\text{-Ha, 8-Hb}}$ = 13, $J_{8\text{-Ha, 2''-H}}$ < 1, $J_{8\text{-Ha, 6''-H}}$ < 1)	6, 1'', 2'', 6''
		8-H$_b$	2.77 ($J_{8\text{-Hb, 2''-H}}$ < 1, $J_{8\text{-Hb, 6''-H}}$ < 1)	6, 1'', 2'', 6''
1'	126.8			

2', 6'	127.4	2'-H, 6'-H	7.04 ($J_{2'\text{-H, }3'\text{-H}}$ = 8, $J_{6'\text{-H, }5'\text{-H}}$ = 8)	7, 1', 4', 6'
3', 5'	112.8	3'-H, 5'-H	6.73	1', 4', 5'
4'	153.6			
		4'-OH		
1''	126.8			
2'', 6''	127.4	2''-H, 6''-H	7.04 ($J_{2''\text{-H, }3''\text{-H}}$ = 8, $J_{6''\text{-H, }5''\text{-H}}$ = 8)	8, 1'', 4'', 6''
3'', 5''	112.8	3''-H, 5''-H	6.73	1'', 4'', 5''
4''	153.6			
		4''-OH		

Table C.5: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **2** in methanol-d_4. Chemical shifts were referenced to δ(C<u>H</u>D_2OD) = 3.31 ppm and δ(^{13}<u>C</u>HD_2OD) = 49.0 ppm.^{13}C chemical shifts were determined from HMBC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.

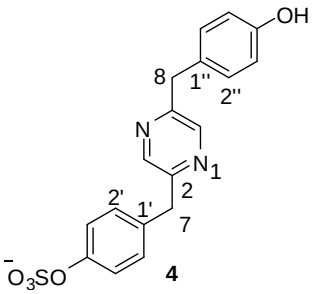


2

Position	**δ_C**	**Proton**	**δ_H (J_{HH}[Hz])**	**HMBC**
1		1-NH		
2	56.6	2-H	3.16 ($J_{2\text{-H, 3-Ha}}$ = 12, $J_{2\text{-H, 3-Hb}}$ = 3, $J_{2\text{-H, 7-Ha}}$ = 6, $J_{2\text{-H, 7-Hb}}$ = 8)	3, 7, 1'
3	49.1	3-H_a	2.69 ($J_{3\text{-Ha, 3-Hb}}$ = 12)	2, 7
		3-H_b	3.12	2
4		4-NH		
5	56.6	5-H	3.16 ($J_{5\text{-H, 6-Ha}}$ = 12, $J_{5\text{-H, 6-Hb}}$ = 3, $J_{5\text{-H, 8-Ha}}$ = 6, $J_{5\text{-H, 8-Hb}}$ = 8)	6, 8, 1''
6	49.1	6-H_a	2.69 ($J_{6\text{-Ha, 6-Hb}}$ = 12)	5, 8
		6-H_b	3.12	5
7	38.3	7-H_a	2.67 ($J_{7\text{-Ha, 7-Hb}}$ = 14, $J_{7\text{-Ha, 2'-H}}$ < 1, $J_{7\text{-Ha, 6'-H}}$ < 1)	2, 3, 1', 2', 6'
		7-H_b	2.76 ($J_{7\text{-Hb, 2'-H}}$ < 1, $J_{7\text{-Hb, 6'-H}}$ < 1)	2, 3, 1', 2', 6'
8	38.3	8-H_a	2.67 ($J_{8\text{-Ha, 8-Hb}}$ = 14, $J_{8\text{-Ha, 2''-H}}$ < 1, $J_{8\text{-Ha, 6''-H}}$ < 1)	5, 6, 1'', 2'', 6''
		8-H_b	2.76 ($J_{8\text{-Hb, 2''-H}}$ < 1, $J_{8\text{-Hb, 6''-H}}$ < 1)	5, 6, 1'', 2'', 6''

1'	127.3			
2', 6'	131.0	2'-H, 6'-H	7.06 ($J_{2'\text{-H, }3'\text{-H}}$ = 8, $J_{6'\text{-H, }5'\text{-H}}$ = 8)	7, 2', 4', 6'
3', 5'	116.5	3'-H, 5'-H	6.75	1', 3', 4', 5'
4'	157.4			
		4'-OH		
1''	127.3			
2'', 6''	131.0	2''-H, 6''-H	7.06 ($J_{2''\text{-H, }3''\text{-H}}$ = 8, $J_{6''\text{-H, }5''\text{-H}}$ = 8)	8, 2'', 4'', 6''
3'', 5''	116.5	3''-H, 5''-H	6.75	1'', 3'', 4'', 5''
4''	157.4			
		4''-OH		

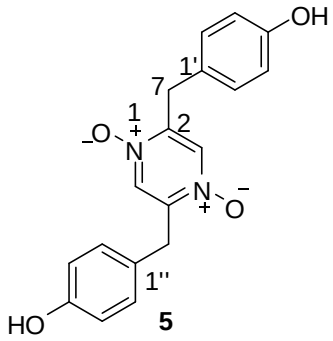
Table C.6: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **4** in methanol-d_4. Chemical shifts were referenced to δ(C<u>H</u>D_2OD) = 3.31 ppm and δ(^{13}<u>C</u>HD_2OD) = 49.0 ppm.^{13}C chemical shifts were determined from HMBC and HSQC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



4

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC
1				
2	154.8			
3	144.1	3-H	8.40 ($J_{3\text{-H, }7\text{-H2}} < 1$)	2, 5
4				
5	155.6			
6	144.3	5-H	8.36 ($J_{6\text{-H, }8\text{-H}} < 1$)	2, 5
7	40.7	7-H_2	4.11 ($J_{7\text{-H2,8-H2}} < 1$, $J_{7\text{-H2,2'-H}} = J_{7\text{-H2,6'-H}} < 1$)	2, 3, 1', 2', 6'
8	40.6	8-H_2	4.02 ($J_{8\text{-H2,2''-H}} = J_{8\text{-H2,6''-H}} < 1$)	5, 6, 1'', 2'', 6''
1'	135.9			
2', 6'	130.2	2'-H, 6'-H	7.23	7, 1', 2', 4', 6'
3', 5'	122.3	3'-H, 5'-H	7.23	Not distinguisable from 2'-H and 6'-H
4'	152.4			
		4'-OSO_3^-		
1''	130.0			
2'', 6''	130.6	2'-H, 6'-H	7.06 ($J_{2''\text{-H,}3''\text{-H}} = 9$)	8, 4''
3'', 5''	116.0	3'-H, 5'-H	6.7	2'', 4'', 6''
4''	156.7			
		4''-OH		

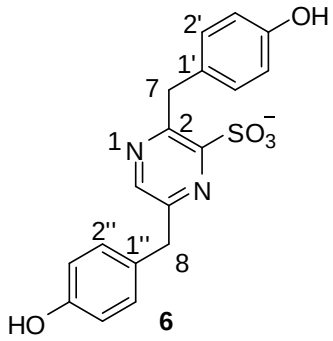
Table C.7: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **5** in DMSO-d_6. Chemical shifts were referenced to δ(C<u>H</u>D_2SOCD_3) = 2.50 ppm and δ(^{13}<u>C</u>HD_2SOCD_3) = 39.5 ppm.^{13}C chemical shifts were determined from HMBC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



5

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC
1				
2	161.9			
3	132.5	3-H	9.04	2, 6
4				
5	161.9			
6	132.5	3-H	9.04	3, 5
7	39.4	7-H_2	4.18	2, 3, 1', 2', 6'
8	39.4	8-H_2	4.18	5, 6, 1'', 2'', 6''
1'	126.4			
2', 6'	129.4	2'-H, 6'-H	6.97 ($J_{2'\text{-H, }3'\text{-H}} = J_{5'\text{-H, }6'\text{-H}} = 9$)	7, 2', 4', 6'
3', 5'	114.7	3'-H, 5'-H	6.70	1', 3', 5'
4'	155.7			
		4'-OH	9.38	3', 4', 5'
1''	126.4			
2'', 6''	129.4	2''-H, 6''-H	6.97 ($J_{2''\text{-H, }3''\text{-H}} = J_{5''\text{-H, }6''\text{-H}} = 9$)	8, 2'', 4'', 6''
3'', 5''	114.7	3''-H, 5''-H	6.70	1'', 3'', 5''
4''	155.7			
		4''-OH	9.38	3'', 4'', 5''

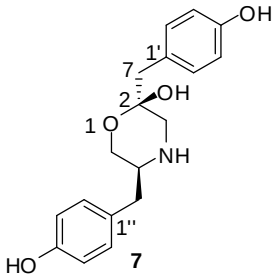
Table C.8: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **6** in DMSO-d_6. Chemical shifts were referenced to δ(C<u>H</u>D_2SOCD_3) = 2.50 ppm and δ(^{13}<u>C</u>HD_2SOCD_3) = 39.5 ppm.^{13}C chemical shifts were determined from HMBC and HMQC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



6

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC
1				
2	150.2			
3	154.3	3-H		
4				
5	150.5			
6	141.9		8.30	2, 3, 5
7	37.1	7-H_2	4.38	2, 3, 2', 6'
8	38.4		3.94	5, 6, 1'', 2'', 6''
1'	129.1			
2', 6'	129.4	2'-H, 6'-H	7.11 ($J_{2'\text{-H, }3'\text{-H}} = J_{5'\text{-H, }6'\text{-H}} = 8$)	7, 2', 4', 6'
3', 5'	113.6	3'-H, 5'-H	6.58	1', 3', 4', 5'
4'	154.2			
		4'-OH	9.05	4', 3', 5'
1''	127.7			
2'', 6''	128.7	2''-H, 6''-H	7.04 ($J_{2''\text{-H, }3''\text{-H}} = J_{5''\text{-H, }6''\text{-H}} = 8$)	8, 2'', 4'', 6''
3'', 5''	114.1	3''-H, 5''-H	6.66	1'', 3'', 4'', 5''
4''	154.7			
		4''-OH	9.24	4'', 3'', 5''

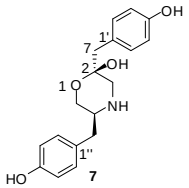
Table C.9: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **afla#7** in methanol-d_4. Chemical shifts were referenced to δ(C<u>H</u>D$_2$OD) = 3.31 ppm and δ(^{13}<u>C</u>HD$_2$OD) = 49.0 ppm.^{13}C chemical shifts were determined from HMBC and HSQC spectra. ^{1}H, ^{1}H-*J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



7

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC
1				
2	93.8			
		2-OH		
3	50.1	3-H$_a$	2.93 ($J_{3\text{-Ha, 3-Hb}}$ = 12)	5
		3-H$_b$	2.89	2, 5
4		4-NH		
5	56.5	5-H	3.32 ($J_{5\text{-H, 6-Ha}}$ = 4, $J_{5\text{-H, 6-Hb}}$ = 11, $J_{5\text{-H, 8-H2}}$ = 8)	3
6	61.5	6-H$_a$	3.66 ($J_{6\text{-Ha, 6-Hb}}$ = 13)	2
		6-H$_b$		2
7	45.4	7-H$_2$	2.88 ($J_{7\text{-Ha, 2'-H}} = J_{7\text{-Ha, 6'-H}} < 1$)	2, 1', 2', 6'
8	34.9	8-H$_2$	2.74 ($J_{8\text{-Ha, 2''-H}} < 1$, $J_{8\text{-Ha, 6''-H}} < 1$)	1''
1'	126.2			
2', 6'	132.2	2'-H, 6'-H	7.07 ($J_{2'\text{-H, 3'-H}} = J_{6'\text{-H, 5'-H}}$ = 9)	2', 4', 6'
3', 5'	115.7	3'-H, 5'-H		1', 3', 5'
4'	157.1			
		4'-OH		
1''	125.8			
2'', 6''	130.8	2''-H, 6''-H	7.05 ($J_{2''\text{-H, 3''-H}} = J_{6''\text{-H, 5''-H}}$ = 9)	8, 1'', 2'', 4'', 6''
3'', 5''	116.4	3''-H, 5''-H		3'', 4'', 1''
4''	157.5			
		4''-OH		

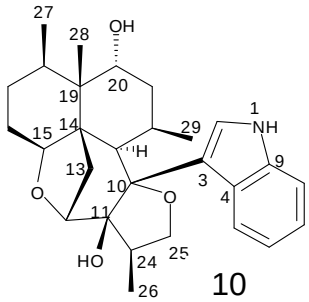
Table C.10: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **7** in DMSO-d_6. Chemical shifts were referenced to δ($C\underline{H}D_2SOCD_3$) = 2.50 ppm and δ($^{13}\underline{C}HD_2SOCD_3$) = 39.5 ppm.^{13}C chemical shifts were determined from HMBC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. Key ROESY correlations were observed using a mixing time of 0.2 s. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



7

Position	δ_C	Proton	δ_H (J_{HH}[Hz])	HMBC	ROESY
1					
2	92.7				
		2-OH	6.72	2, 3	3-H_a, 6-H_b
3	48.2	3-H_a	3.33 ($J_{3\text{-Ha, 3-Hb}}$ = 13)		
		3-H_b	4.10		
4		4-NH	8.08		
5	54.7	5-H	3.23 ($J_{5\text{-H, 6-Ha}}$ = 3, $J_{5\text{-H, 6-Hb}}$ = 11, $J_{5\text{-H, 8-Ha}}$ = 6, $J_{5\text{-H, 8-Ha}}$ = 8, $J_{5\text{-H, 8-Hb}}$ = 6)		
6	60.1	6-H_a	3.54 ($J_{6\text{-Ha, 6-Hb}}$ = 13)	2	
		6-H_b	3.90	2, 5	
7	43.7	7-H_a	2.82 ($J_{7\text{-Ha, 2'-H}}$ = $J_{7\text{-Ha, 6'-H}}$ < 1)	3, 1', 2', 6'	
		7-H_b	2.80	3, 1', 2', 6'	
8	33.3	8-H_a	2.64 ($J_{8\text{-Ha, 8-Hb}}$ = 13, $J_{8\text{-Ha, 2''-H}}$ = $J_{8\text{-Ha, 6''-H}}$ < 1)	5, 6, 1'', 2'', 6''	
		8-H_b	2.73	5, 6, 1'', 2'', 6''	
1'	124.8				
2', 6'	131.0	2'-H, 6'-H	7.03 ($J_{2'\text{-H, 3'-H}}$ = $J_{6'\text{-H, 5'-H}}$ = 9)	7	
3', 5'	114.6	3'-H, 5'-H	6.70	3', 4', 5'	
4'	155.8				
		4'-OH	9.36	3', 4', 5'	
1''	124.9				
2'', 6''	129.9	2''-H, 6''-H	7.04 ($J_{2''\text{-H, 3''-H}}$ = $J_{6''\text{-H, 5''-H}}$ = 9)	8	
3'', 5''	115.1	3''-H, 5''-H	6.67	1'', 3'', 4'', 5''	
4''	156.0				
		4''-OH	9.27	3'', 4'', 5''	

Table C.11: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compound **10** in methanol-d_4 (DMSO-d_6, square brackets). Spectra acquired for **10** dissolved in methanol-d_4 were referenced to δ(C<u>H</u>D_2OD) = 3.31 and δ(^{13}<u>C</u>HD_2OD) = 49.0 ppm. Spectra acquired for **10** dissolved in DMSO-d_6 were referenced to δ(C<u>H</u>D_2SOCD_3) = 2.50 ppm and δ(^{13}<u>C</u>HD_2SOCD_3) = 39.5 ppm.^{13}C chemical shifts were determined from gHMBCAD and HSQC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra ($\delta H_{DMSO\text{-}d6}$ chemical shifts and coupling constants in square brackets). ROESY correlations were observed using a mixing time of 0.2 s. Key HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom, weak correlations are indicated by (w).

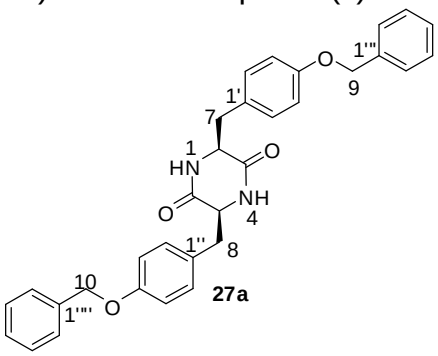


10

Position	δ_C	Proton	δ_H (methanol-d_4) [δ_H (DMSO-d_6) (J_{HH}[Hz])]	HMBC	ROESY
1		1-NH	[10.6 ($J_{1\text{-NH, 2-H}}$ = 2)]		
2	127.9	2-H	7.90 [7.82]	3, 4, 9	11-OH, 25-H_a(w), 26-H_3, 29-H_3
3	116.8				
4	128.6				
5	121.9	5-H	7.86 [7.79 ($J_{5\text{-H, 6-H}}$ = 8)]	9	22-H, 23-H
6	119.1	6-H	6.93 [6.88 ($J_{6\text{-H, 7-H}}$ = 8)]	4	
7	121.5	7-H	7.03 [6.99 ($J_{7\text{-H, 8-H}}$ = 8)]	9	
8	111.5		7.30 [7.28]		
9	136.9				
10	88.8				
11	81.4				12-H, 13-H, 26-H_3
		11-OH	[4.79]		
12	86.3	12-H	3.93 [3.84 ($J_{12\text{-H, 13-Ha}}$ = 6)]	10, 11, 14, 15	24-H
13	29.6	13-H_a	1.83 [1.67 ($J_{13\text{-Ha, 13-Hb}}$ = 12)]	11, 12	28-H_3
		13-H_b	2.36 [2.18]		20-H, 29-H_3
14	50.6				
15	85.5	15-H	4.75 [4.55 ($J_{15\text{-H, 16-Ha}}$ = 14, $J_{15\text{-H, 16-Hb}}$ = 6)]	23	16-H_b, 18-H, 23-H
16	27.1	16-H_a	1.63 [1.49 ($J_{16\text{-Ha, 17-Ha}}$ = 9)]		

		16-H_b	1.73 [1.58]	14	18-H
17	29.3	17-H_a	1.34 [1.17 ($J_{17\text{-Ha, }17\text{-Hb}}$ = 13)]		
		17-H_b	1.34 [1.61]		
18	30.3	18-H	2.12 [1.99 ($J_{18\text{-H, }27\text{-CH3}}$ = 7)]		
19	45.5				
20	75.4	20-H	3.78 [3.60 ($J_{20\text{-H, }20\text{-OH}}$ = 7, $J_{20\text{-H, }21\text{-Ha}}$ = 4, $J_{20\text{-H, }21\text{-Hb}}$ = 13)]		21-H_a, 28-H_3, 29-H_3
		20-OH	[4.11]		
21	38.1	21-H_a	1.25 [1.14 ($J_{21\text{-Ha, }21\text{-Hb}}$ = 13, $J_{21\text{-Ha, }22\text{-H}}$ = 13)]		22-H, 29-CH_3
		21-H_b	1.67 [1.48 ($J_{21\text{-Hb, }22\text{-H}}$ = 5)]	19	
22	31.7	22-H	1.54 [1.39 ($J_{22\text{-H, }23\text{-H = 5}}$, $J_{22\text{-H, }29\text{-CH3}}$ = 7)]		23-H,
23	51.6	23-H	2.87 [2.73]	10, 11, 14, 15	
24	41.5	24-H	2.40 [2.31 ($J_{24\text{-H, }25\text{-Ha}}$ = 8, $J_{24\text{-H, }25\text{-Hb}}$ = 9)]		12-H, 25-H_b, 26-H_3
25	72.6	25-H_a	2.77 [2.59 ($J_{25\text{-Ha, }25\text{-Hb}}$ = 9)]		26-H_3
		25-H_b	3.90 [3.79]	10, 11	
26	15.1	26-H_3	0.88 [0.81]	11	25-H_a
27	18.7	27-H_3	1.02 [0.94]	17, 19	13-H_a, 20-H, 20-OH, 21-H_b, 28-H_3
28	19.3	28-H_3	1.18 [1.07]	14, 18, 19, 20	
29	16.5	29-H_3	1.03 [0.94]		13-H_b, 18-H, 20-OH(w), 20-H, 21-H_a

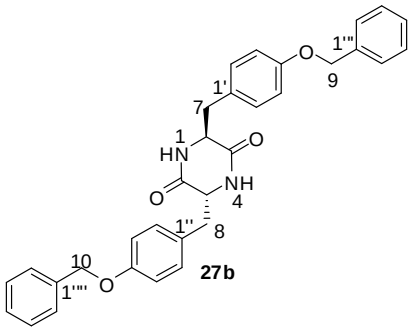
Table S12: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **27a** in chloroform-*d*. Chemical shifts were referenced to δ($C\underline{H}Cl_3$) = 7.26 ppm and δ($^{13}\underline{C}HCl_3$) = 77.2 ppm.^{13}C chemical shifts were determined from HMBC spectra. $^{1}H,^{1}H$ *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.



Position	δ_C	**Proton**	δ_H (J_{HH}[Hz])	**HMBC**
1		1-NH	5.71 ($J_{1\text{-NH, 2-H}}$ = 3)	2, 3
2	56.0	2-H	4.09 ($J_{2\text{-H, 7-Ha}}$ = 8, $J_{2\text{-H, 7-Hb}}$ = 5)	3, 1'
3	165.8			
4		4-NH	5.71 ($J_{4\text{-NH, 5-H}}$ = 3)	5, 6
5	56.0		4.09 ($J_{5\text{-H, 8-Ha}}$ = 8, $J_{5\text{-H, 8-Hb}}$ = 5)	6, 1''
6	165.8			
7	39.1	7-H_a	2.31 ($J_{7\text{-Ha, 7-Hb}}$ = 14, $J_{7\text{-Ha, 2'-H}}$ < 1, $J_{7\text{-Ha, 6'-H}}$ < 1)	2, 3, 1', 2', 6,'
		7-H_b	3.03 ($J_{7\text{-Hb, 2'-H}}$ < 1, $J_{7\text{-Hb, 6'-H}}$ < 1)	2, 3, 1', 2', 6,'
8	39.1	8-H_a	2.31 ($J_{8\text{-Ha, 8-Hb}}$ = 14, $J_{8\text{-Ha, 2''-H}}$ < 1, $J_{8\text{-Ha, 6''-H}}$ < 1)	5, 6, 1'', 2'', 6,''
		8-H_b	3.03 ($J_{8\text{-Hb, 2''-H}}$ < 1, $J_{8\text{-Hb, 6''-H}}$ < 1)	5, 6, 1'', 2'', 6,''
9	69.3	9-H	5.05 ($J_{9\text{-H, 3'-H}}$ < 1, $J_{9\text{-H, 5'-H}}$ < 1, $J_{9\text{-H, 2'''-H}}$ < 1, $J_{9\text{-H, 6'''-H}}$ < 1)	4', 1''', 2''', 6'''
10	69.3		5.05 ($J_{10\text{-H, 3''-H}}$ < 1, $J_{10\text{-H, 5''-H}}$ < 1, $J_{10\text{-H, 2''''-H}}$ < 1, $J_{10\text{-H, 6''''-H}}$ < 1)	4'', 1'''', 2'''', 6''''
1'	127.0			
2', 6'	130.5	2'-H, 6'-H	7.04 ($J_{2'\text{-H, 3'-H}}$ = 8, $J_{6'\text{-H, 5'-H}}$ = 8)	7, 2', 4', 6'
3', 5'	115.2	3'-H, 5'-H	6.95	1', 3', 4', 5'
4'	157.8			
1''	127.0			

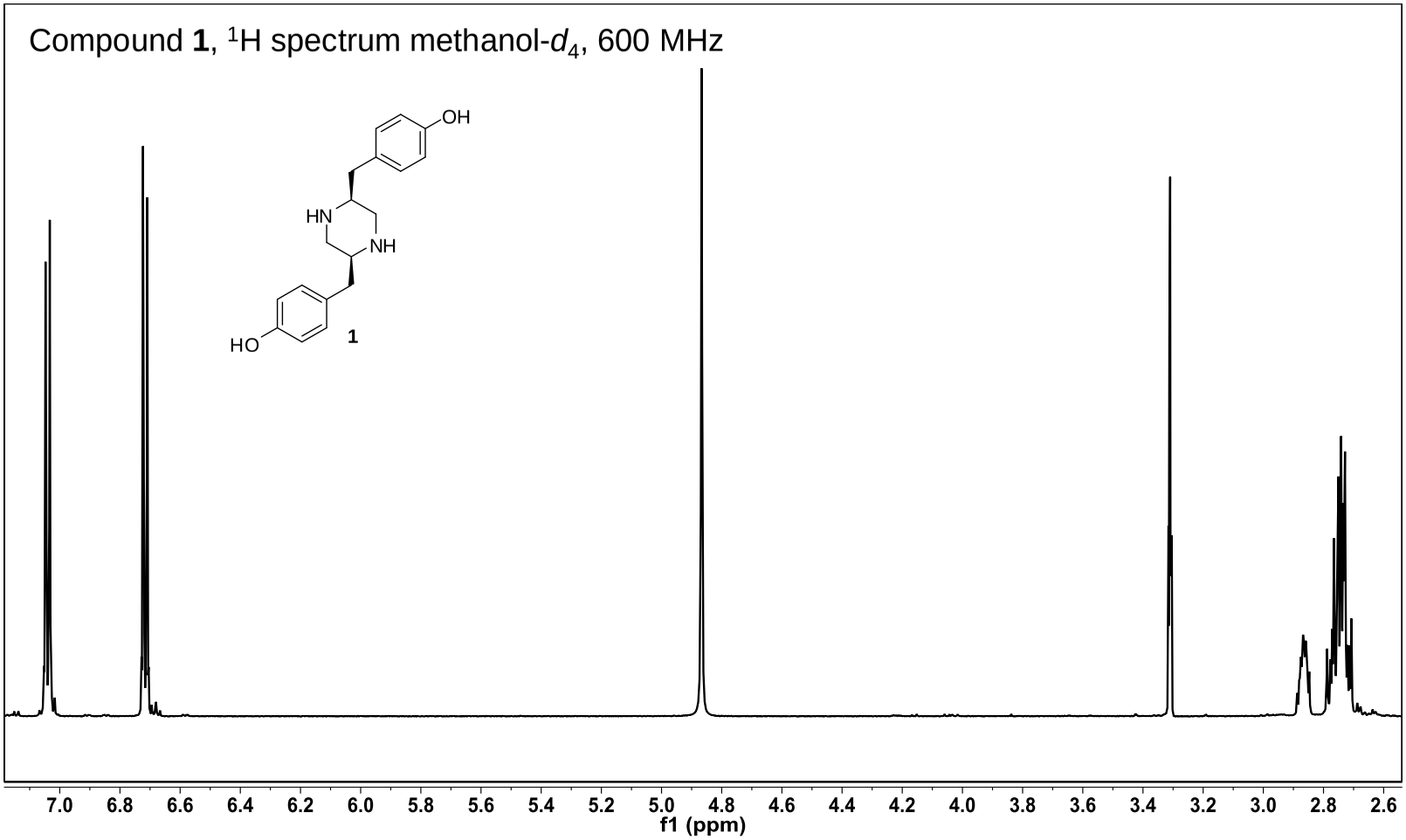
2'', 6''	130.5	2''-H, 6''-H	7.04 ($J_{2''\text{-H, }3''\text{-H}}$ = 8, $J_{6''\text{-H, }5''\text{-H}}$ = 8)	8, 2'', 4'', 6''
3'', 5''	115.2	3''-H, 5''-H	6.95	1'', 3'', 4'', 5''
4''	157.8			
1'''	136.3	1'''-H		
2''', 6'''	126.9	2'''-H, 6'''-H	7.39 ($J_{2'''\text{-H, }3'''\text{-H}}$ = 8, $J_{6'''\text{-H, }5'''\text{-H}}$ = 8)	9, 4'''
3''', 5'''	128.2	3'''-H, 5'''-H	7.33 ($J_{3'''\text{-H, }4'''\text{-H}}$ = 7, $J_{5'''\text{-H, }4'''\text{-H}}$ = 7)	1''', 3''', 5'''
4'''	127.4	4'''-H	7.30	2''', 6'''
1''''	136.3			
2'''', 6''''	126.9	2''''-H, 6''''-H	7.39 ($J_{2''''\text{-H, }3''''\text{-H}}$ = 8, $J_{6''''\text{-H, }5''''\text{-H}}$ = 8)	10, 4''''
3'''', 5''''	128.2	3''''-H, 5''''-H	7.33 ($J_{3''''\text{-H, }4''''\text{-H}}$ = 7, $J_{5''''\text{-H, }4''''\text{-H}}$ = 7)	1'''', 3'''', 5''''
4''''	127.4	4''''-H	7.30	2'''', 6''''

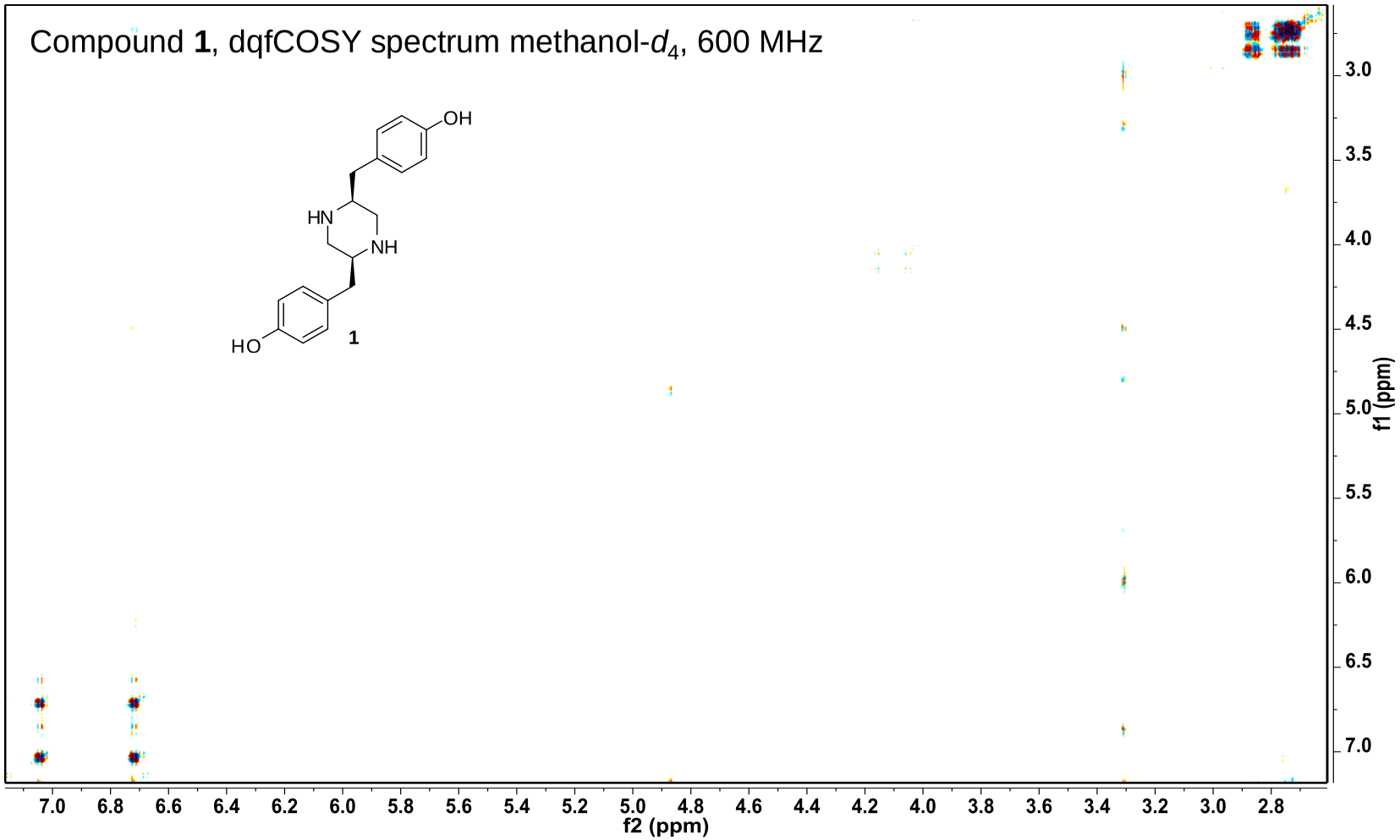
Table S13: ^{1}H (600 MHz) and ^{13}C (151 MHz) NMR spectroscopic data for compounds **27b** in chloroform-*d*. Chemical shifts were referenced to δ(C<u>H</u>Cl$_3$) = 7.26 ppm and δ(^{13}<u>C</u>HCl$_3$) = 77.2 ppm.^{13}C chemical shifts were determined from HMBC and HMQC spectra. ^{1}H,^{1}H *J*-coupling constants were determined from the acquired ^{1}H or dqfCOSY spectra. HMBC correlations (optimized for 6 Hz) are from the proton(s) stated to the indicated ^{13}C atom.

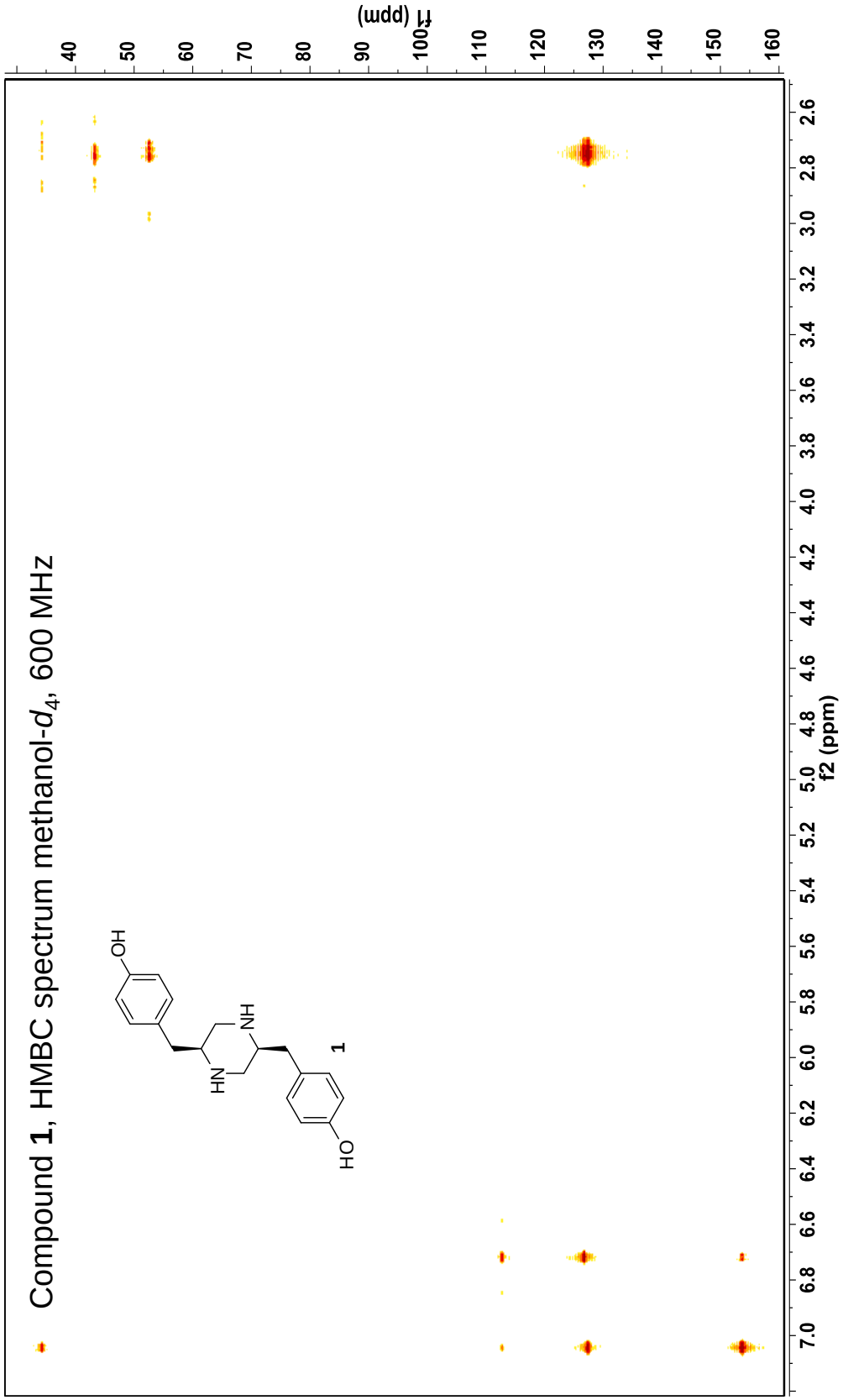


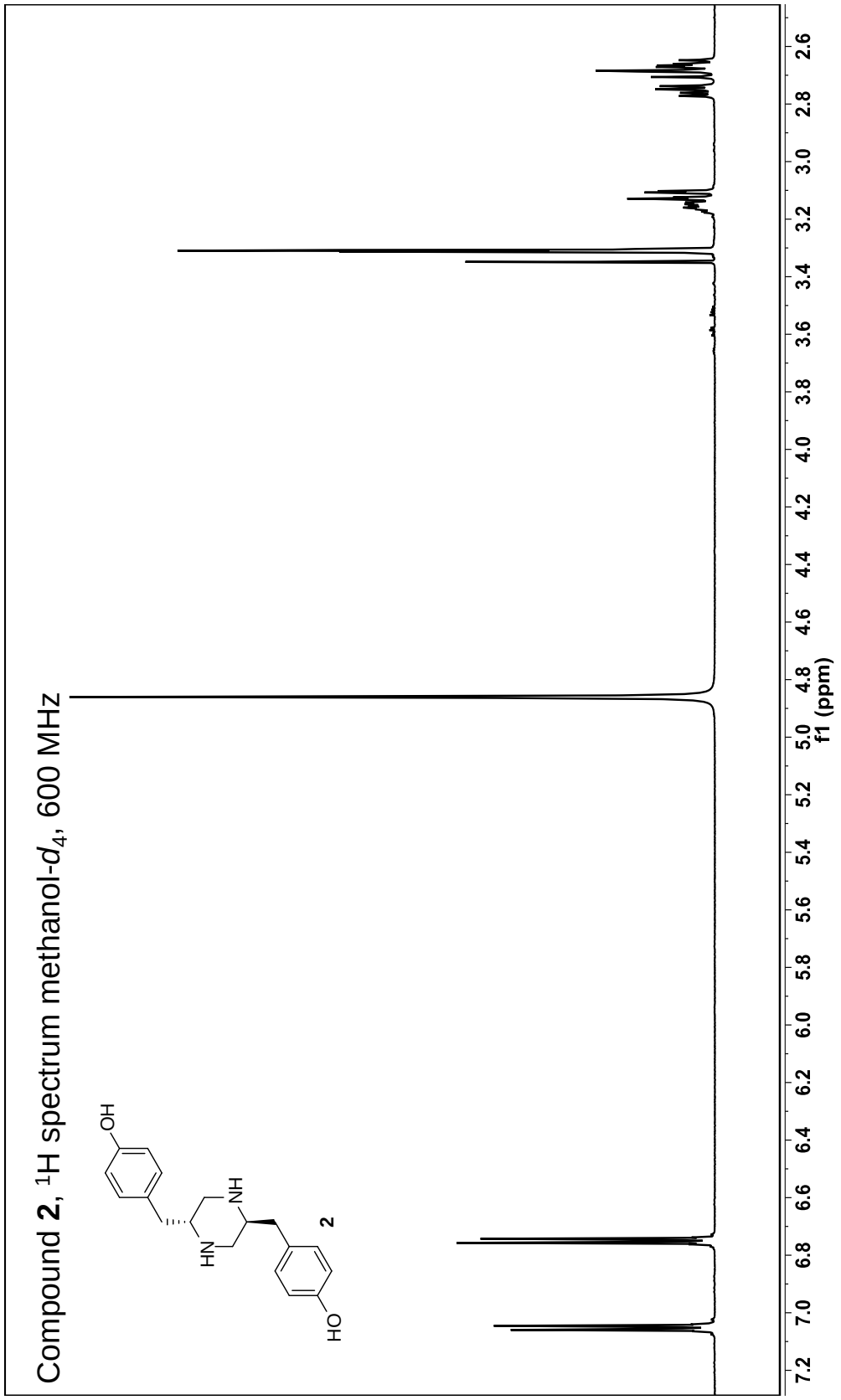
Position	**δ_C**	**Proton**	**δ_H (J_{HH}[Hz])**	**HMBC**
1		1-NH	5.61 ($J_{1\text{-NH, }2\text{-H}}$ = 3)	3
2	55.7	2-H	3.68 ($J_{2\text{-H, }7\text{-Ha}}$ = 8, $J_{2\text{-H, }7\text{-Hb}}$ = 3)	3
3	167.2			
4		4-NH	5.61 ($J_{4\text{-NH, }5\text{-H}}$ = 3)	3
5	55.7		3.68 ($J_{5\text{-H, }8\text{-Ha}}$ = 8, $J_{5\text{-H, }8\text{-Hb}}$ = 3)	3
6	167.2			
7	38.8	7-H_a	2.83 ($J_{7\text{-Ha, }7\text{-Hb}}$ = 15, $J_{7\text{-Ha, }2'\text{-H}}$ < 1, $J_{7\text{-Ha, }6'\text{-H}}$ < 1)	3, 1'
		7-H_b	3.15 ($J_{7\text{-Hb, }2'\text{-H}}$ < 1, $J_{7\text{-Hb, }6'\text{-H}}$ < 1)	3
8	38.8	8-H_a	2.83 ($J_{8\text{-Ha, }8\text{-Hb}}$ = 15, $J_{8\text{-Ha, }2''\text{-H}}$ < 1, $J_{8\text{-Ha, }6''\text{-H}}$ < 1)	6, 1''
		8-H_b	3.15 ($J_{8\text{-Hb, }2''\text{-H}}$ < 1, $J_{8\text{-Hb, }6''\text{-H}}$ < 1)	6
9	70.0	9-H	5.05 ($J_{9\text{-H, }3'\text{-H}}$ < 1, $J_{9\text{-H, }5'\text{-H}}$ < 1, $J_{9\text{-H, }2'''\text{-H}}$ < 1, $J_{9\text{-H, }6'''\text{-H}}$ < 1)	4', 1''', 2''', 6'''
10	70.0		5.05 ($J_{10\text{-H, }3''\text{-H}}$ < 1, $J_{10\text{-H, }5''\text{-H}}$ < 1, $J_{10\text{-H, }2''''\text{-H}}$ < 1, $J_{10\text{-H, }6''''\text{-H}}$ < 1)	4'', 1'''', 2'''', 6''''
1'	126.7			
2', 6'	130.5	2'-H, 6'-H	7.05 ($J_{2'\text{-H, }3'\text{-H}}$ = 9, $J_{6'\text{-H, }5'\text{-H}}$ = 9)	7, 4'
3', 5'	115.3	3'-H, 5'-H	6.92	
4'	158.3			
1''	126.7			

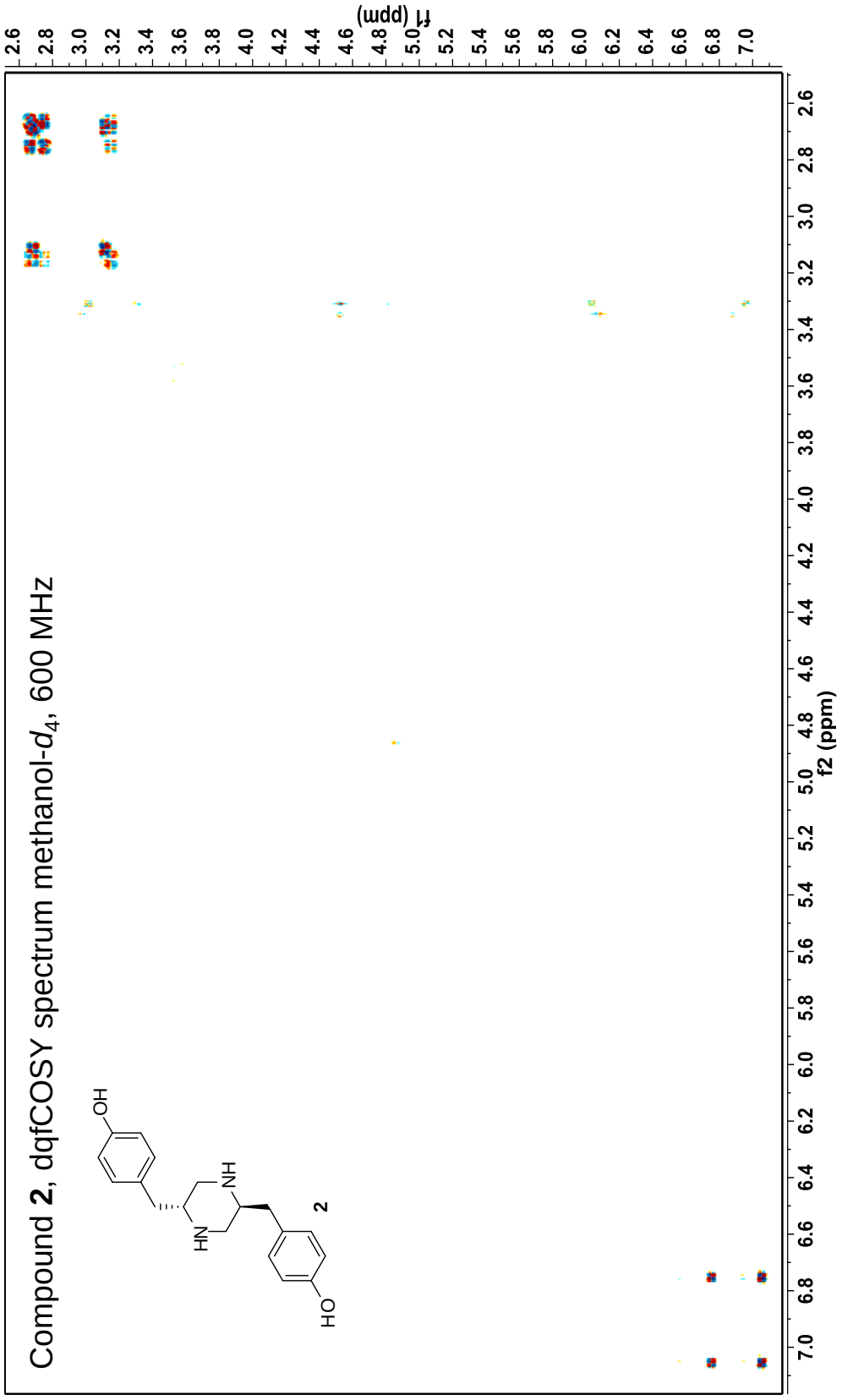
2'', 6''	130.5	2''-H, 6''-H	7.05 ($J_{2''\text{-H, }3''\text{-H}}$ =, $J_{6''\text{-H, }5''\text{-H}}$ =)	8, 4''
3'', 5''	115.3	3''-H, 5''-H	6.92	
4''	158.3			
1'''	136.7	1'''-H		
2''', 6'''	127.3	2'''-H, 6'''-H	7.44 ($J_{2'''\text{-H, }3'''\text{-H}}$ = 9, $J_{6'''\text{-H, }5'''\text{-H}}$ = 9)	
3''', 5'''	128.5	3'''-H, 5'''-H	7.40 ($J_{3'''\text{-H, }4'''\text{-H}}$ = 8, $J_{5'''\text{-H, }4'''\text{-H}}$ = 8)	
4'''	128.2	4'''-H	7.34	
1''''	136.7			
2'''', 6''''	127.3	2''''-H, 6''''-H	7.44 ($J_{2''''\text{-H, }3''''\text{-H}}$ = 9, $J_{6''''\text{-H, }5''''\text{-H}}$ = 9)	
3'''', 5''''	128.5	3''''-H, 5''''-H	7.40 ($J_{3''''\text{-H, }4''''\text{-H}}$ = 8, $J_{5''''\text{-H, }4''''\text{-H}}$ = 8)	
4''''	128.2	4''''-H	7.34	

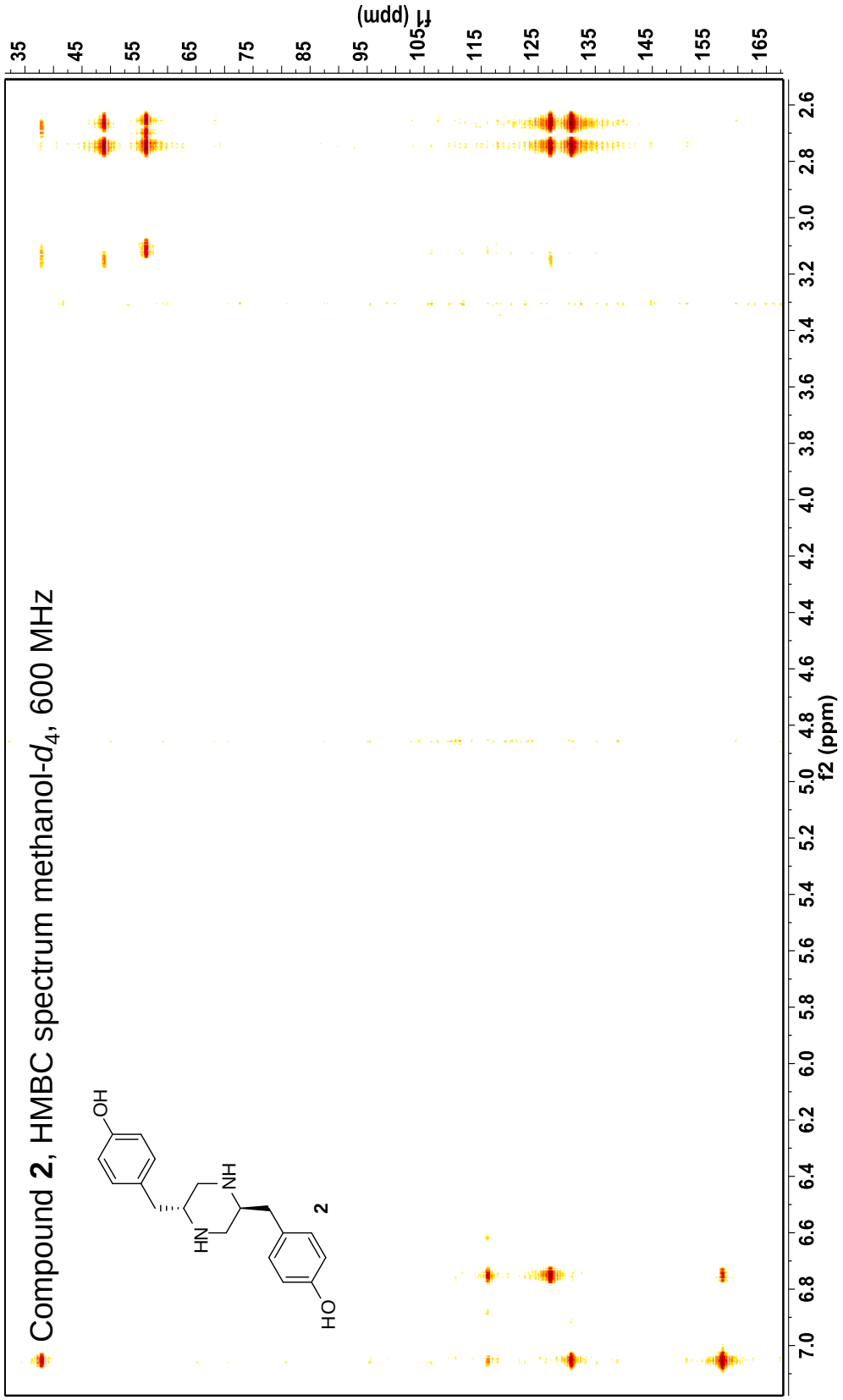

Compound **1**, ^{1}H spectrum methanol-d_4, 600 MHz
OH
HN
NH
HO
1
7.0
6.8
6.6
6.4
6.2
6.0
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
f1 (ppm)

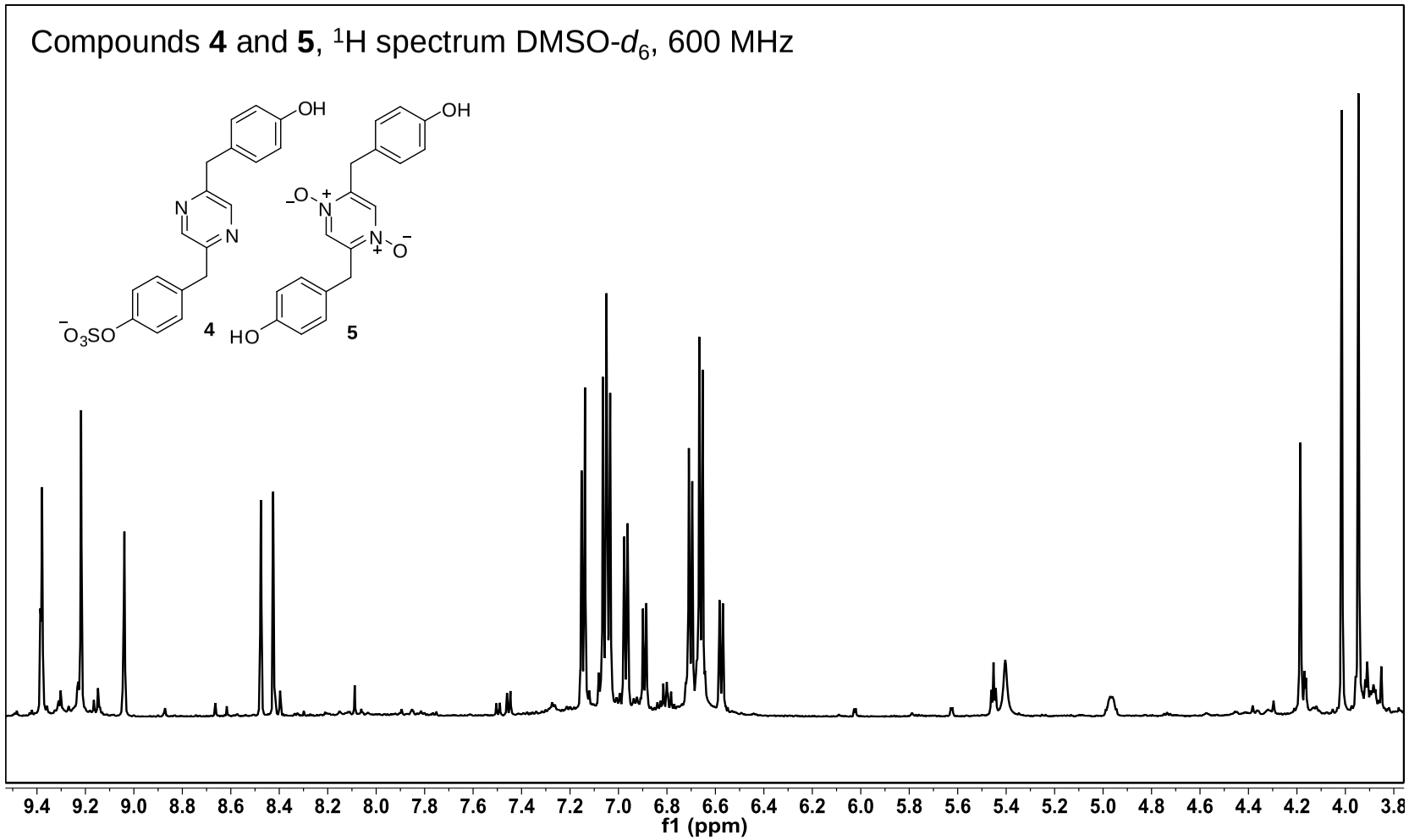

Compound **1**, dqfCOSY spectrum methanol-d_4, 600 MHz
OH
HN
NH
HO
1
f1 (ppm)
f2 (ppm)

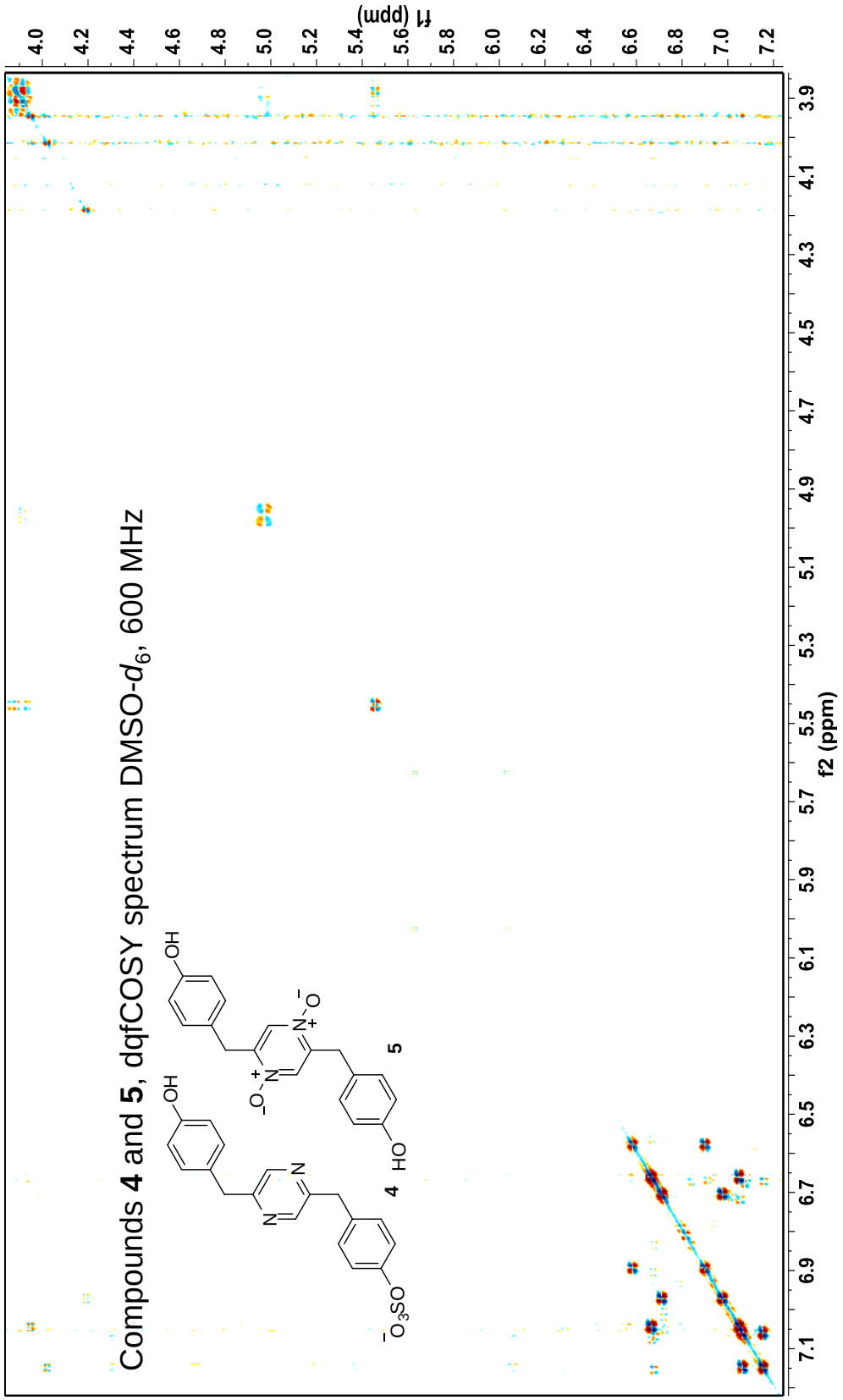

Compound 1, HMBC spectrum methanol-d_4, 600 MHz
f1 (ppm)
f2 (ppm)
OH
NH
NH
OH
1

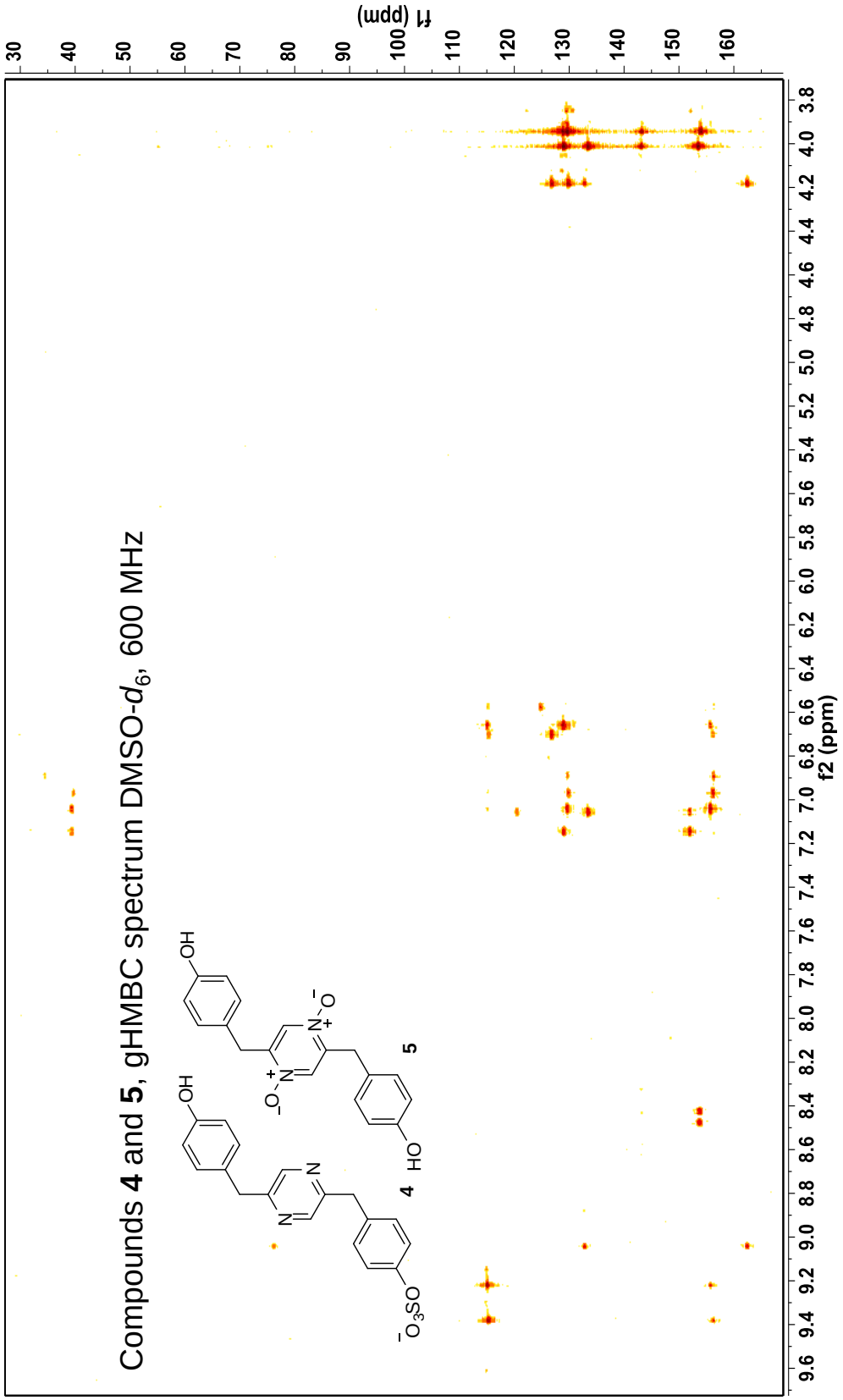

Compound 2, ^{1}H spectrum methanol-d_4, 600 MHz
OH
NH
HN
2
HO
7.2
7.0
6.8
6.6
6.4
6.2
6.0
5.8
5.6
5.4
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
f1 (ppm)

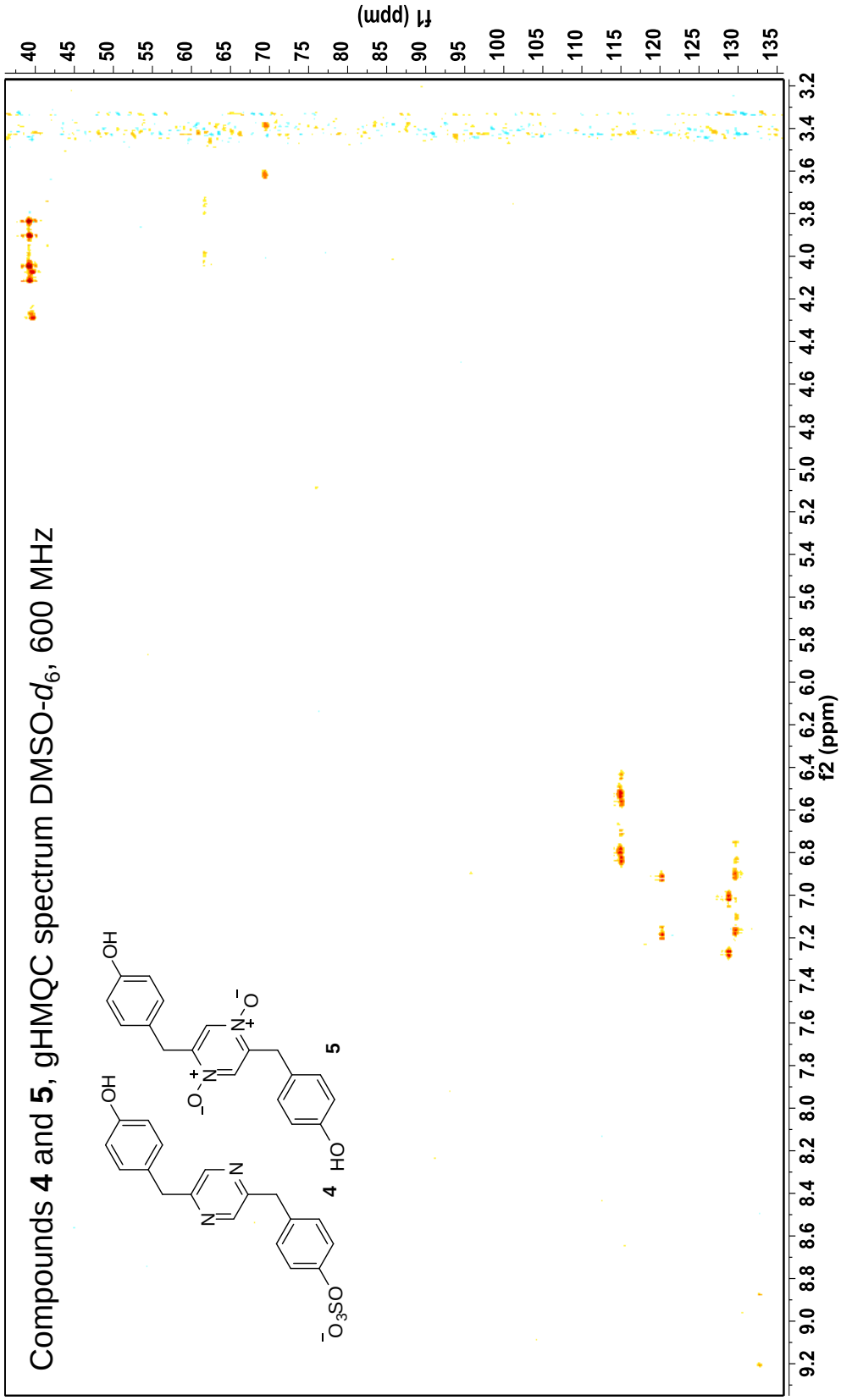

Compound 2, dqfCOSY spectrum methanol-d_4, 600 MHz
f1 (ppm)
f2 (ppm)
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
5.0
5.2
5.4
5.6
5.8
6.0
6.2
6.4
6.6
6.8
7.0
OH
NH
HN
HO
2

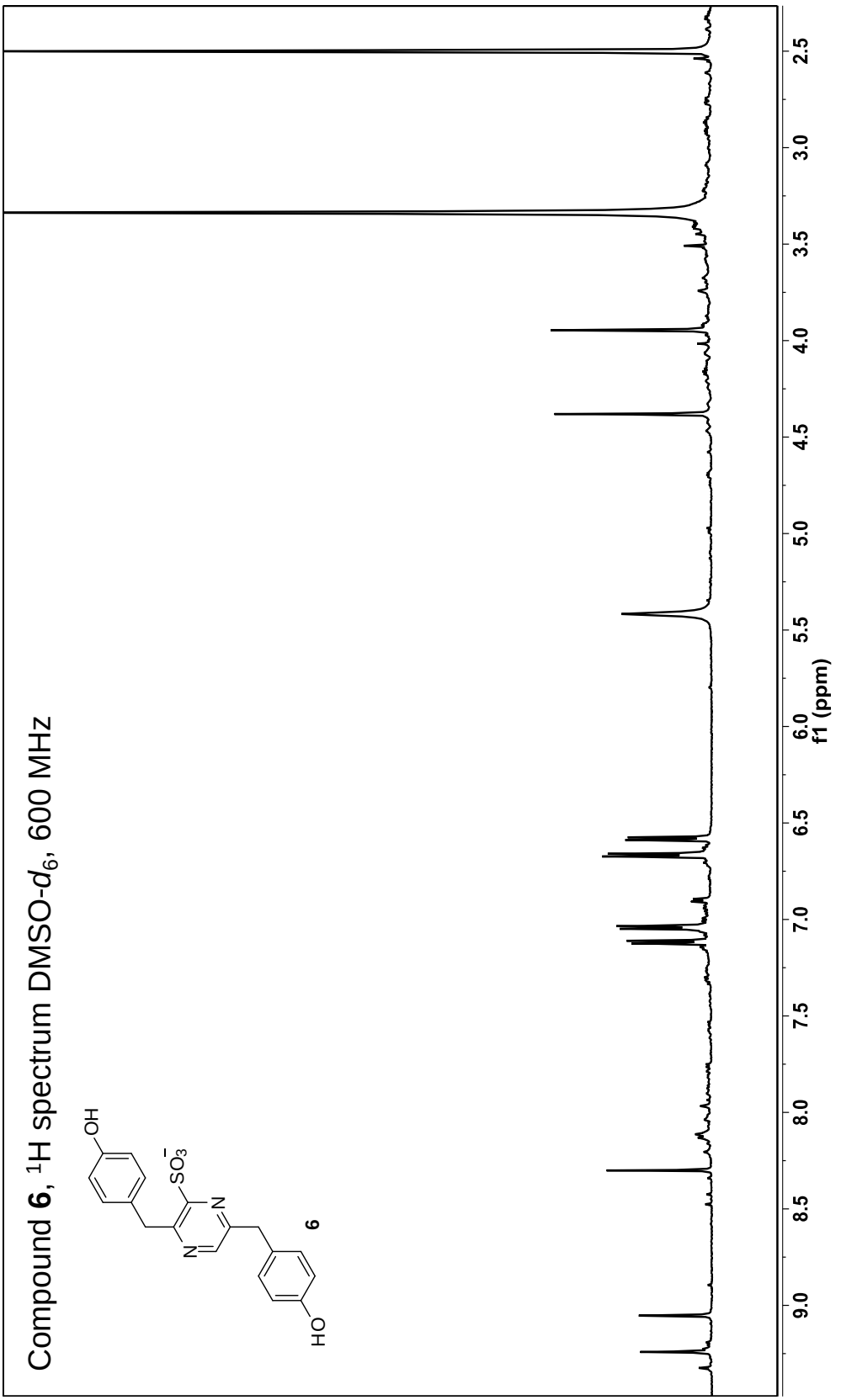

Compound 2, HMBC spectrum methanol-d_4, 600 MHz
f1 (ppm)
f2 (ppm)
OH
HN
NH
HO
2

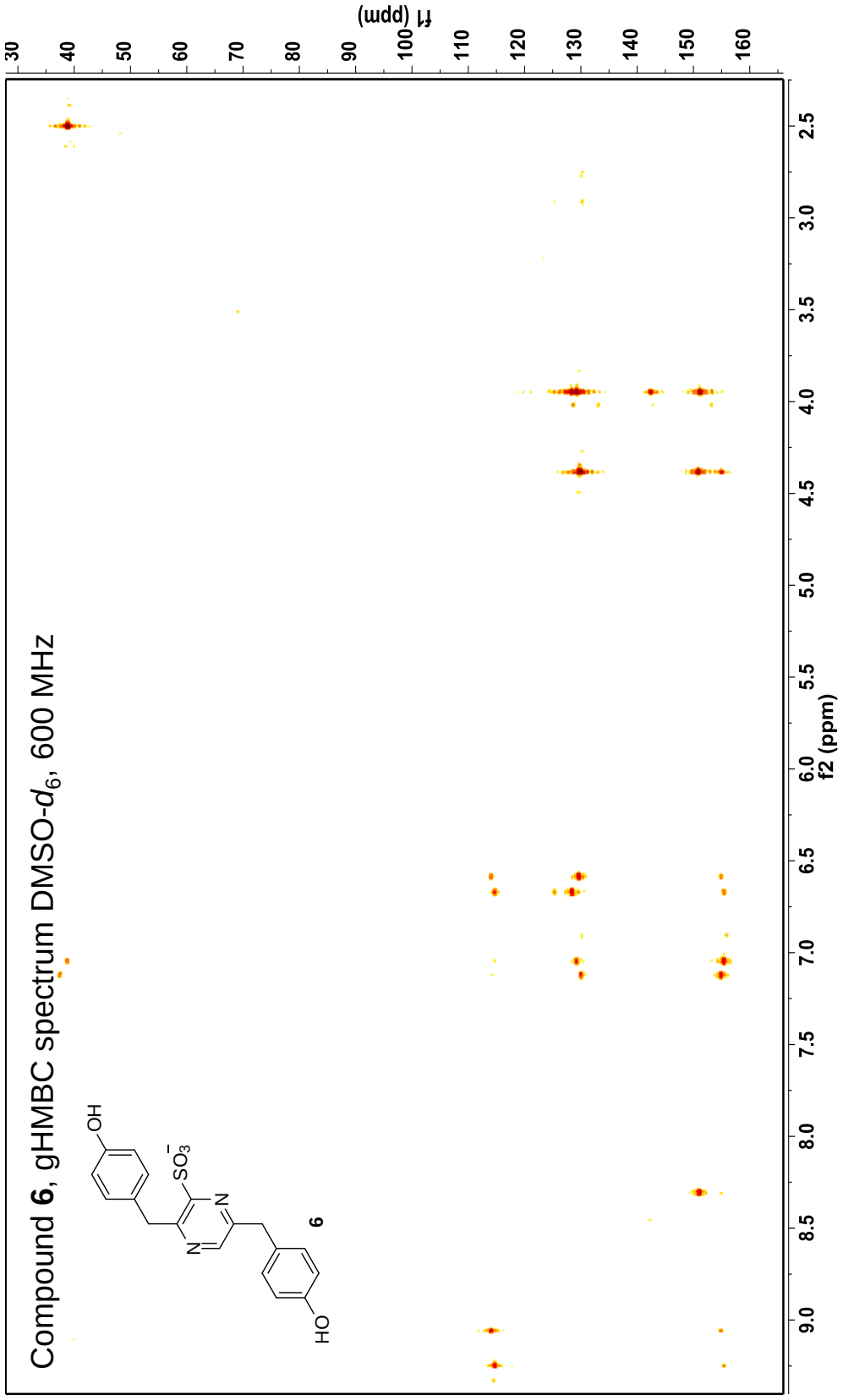

Compounds **4** and **5**, ^{1}H spectrum DMSO-d_6, 600 MHz
OH
N
N
^-O_3SO
4
OH
^-O
N
$^+$
N
$^+$
^-O
HO
5
9.4 9.2 9.0 8.8 8.6 8.4 8.2 8.0 7.8 7.6 7.4 7.2 7.0 6.8 6.6 6.4 6.2 6.0 5.8 5.6 5.4 5.2 5.0 4.8 4.6 4.4 4.2 4.0 3.8
f1 (ppm)


Compounds 4 and 5, dqfCOSY spectrum DMSO-d_6, 600 MHz
f1 (ppm)
f2 (ppm)
4.0
4.2
4.4
4.6
4.8
5.0
5.2
5.4
5.6
5.8
6.0
6.2
6.4
6.6
6.8
7.0
7.2
3.9
4.1
4.3
4.5
4.7
4.9
5.1
5.3
5.5
5.7
5.9
6.1
6.3
6.5
6.7
6.9
7.1
OH
HO
^-O_3SO
4
5

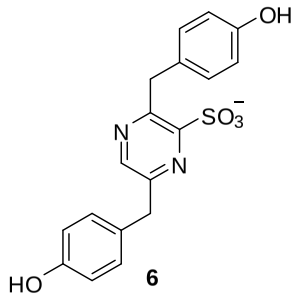

Compounds 4 and 5, gHMBC spectrum DMSO-d6, 600 MHz
f1 (ppm)
f2 (ppm)
4
5


Compounds 4 and 5, gHMQC spectrum DMSO-d_6, 600 MHz
f1 (ppm)
f2 (ppm)
4
5

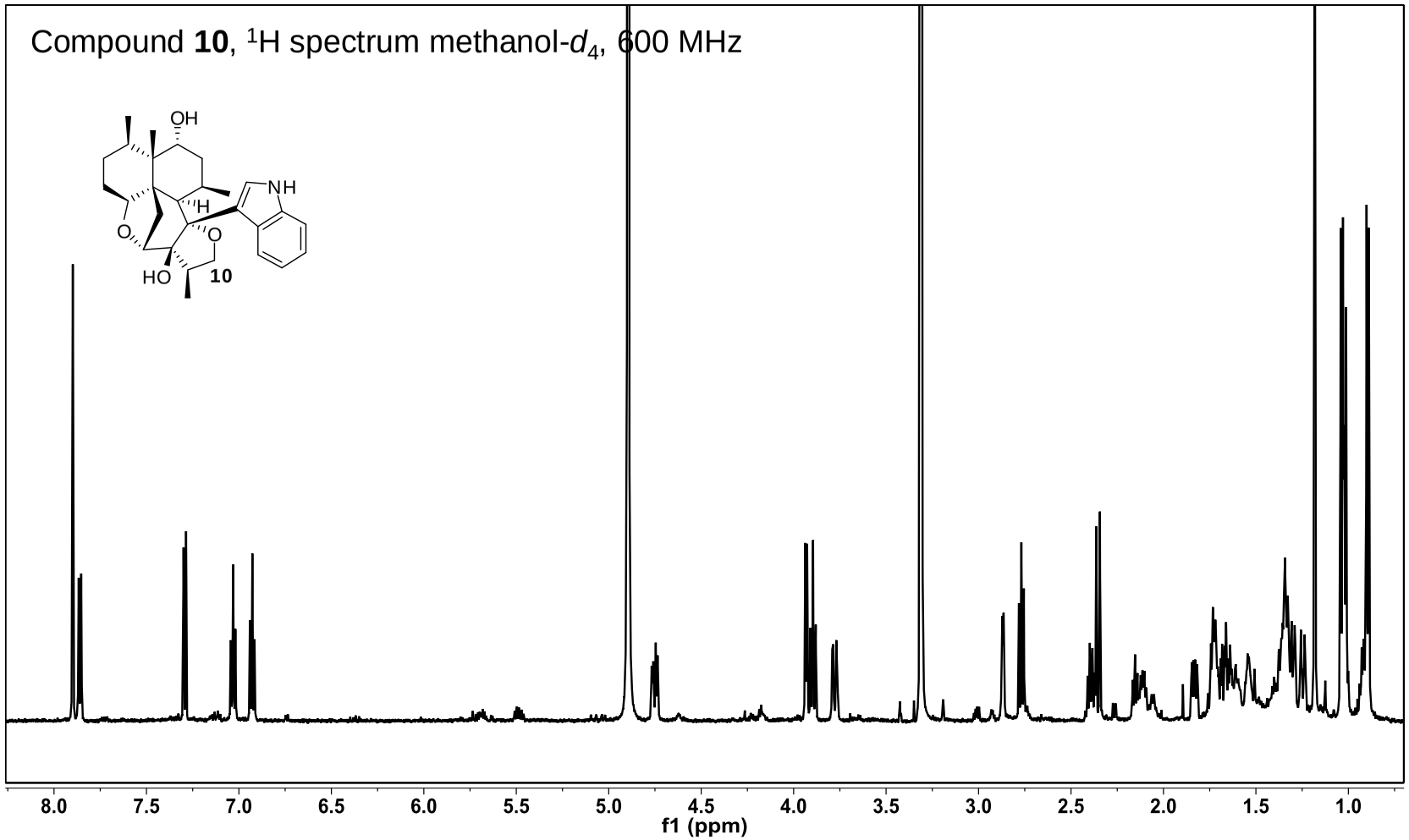

Compound 6, 1H spectrum DMSO-d6, 600 MHz
OH
SO3−
N
N
HO
6
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
f1 (ppm)

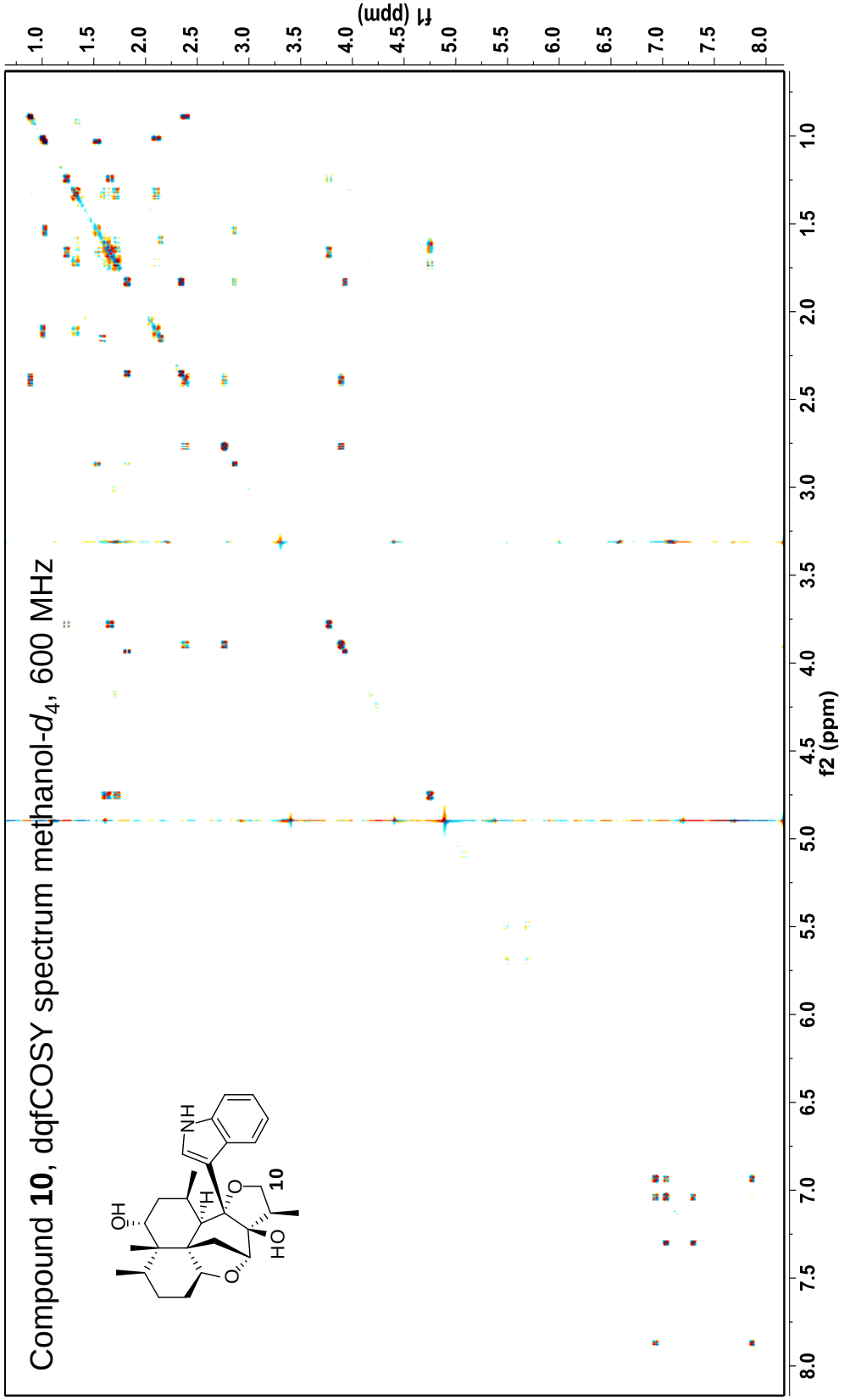

Compound 6, gHMBC spectrum DMSO-d_6, 600 MHz
f1 (ppm)
f2 (ppm)
OH
SO_3^-
N
N
HO
6

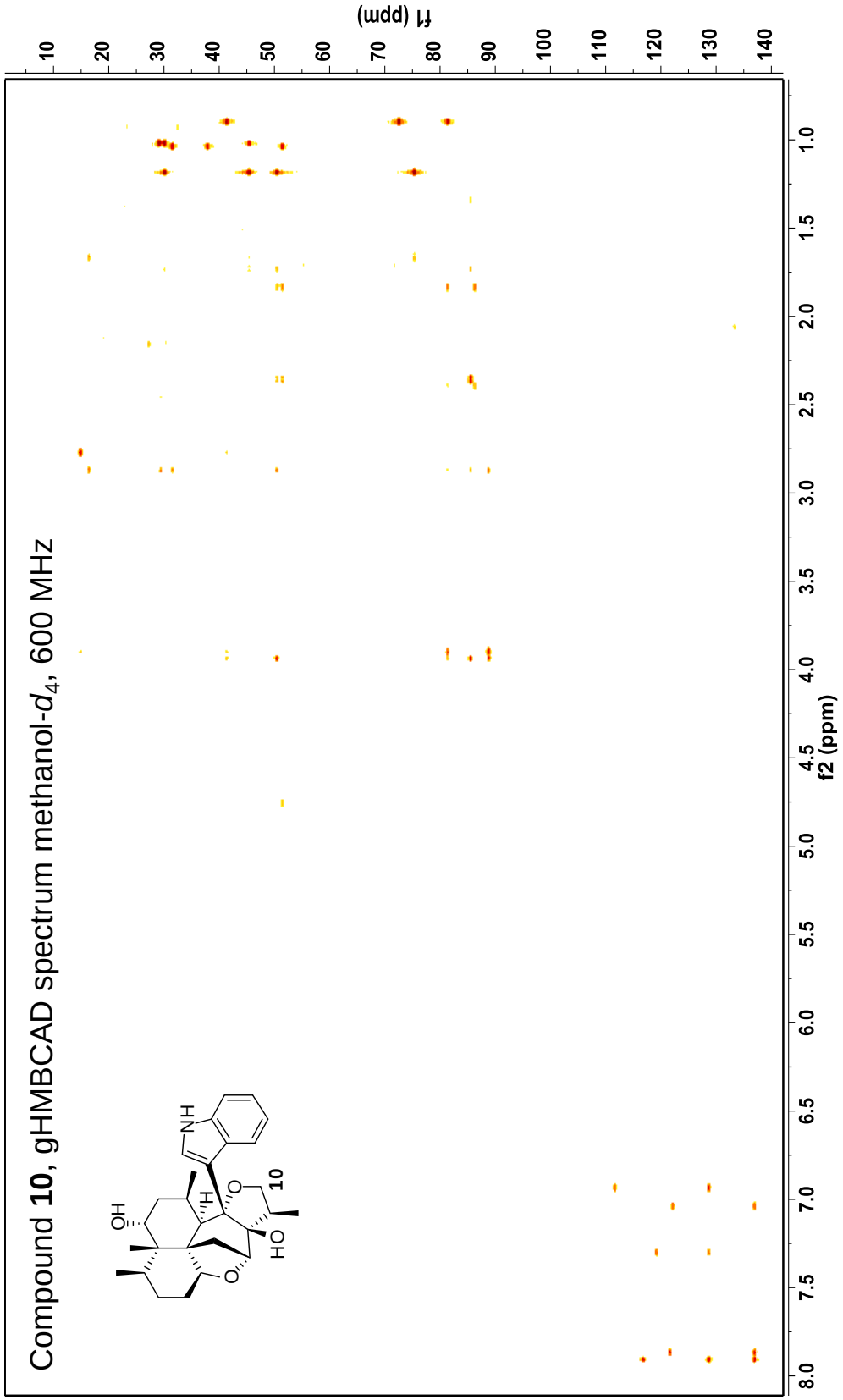
Compound **6**, gHMQC spectrum DMSO-d_6, 600 MHz

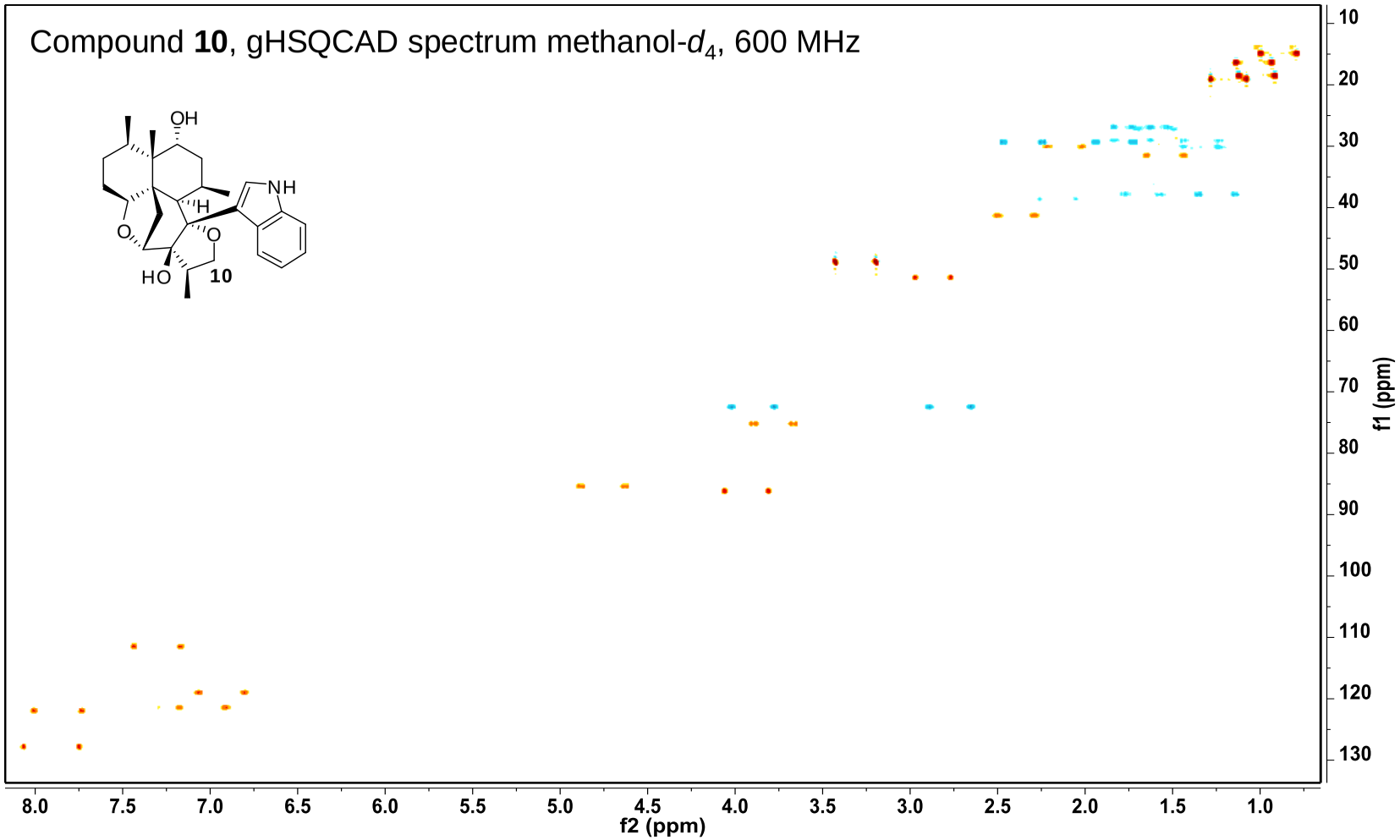


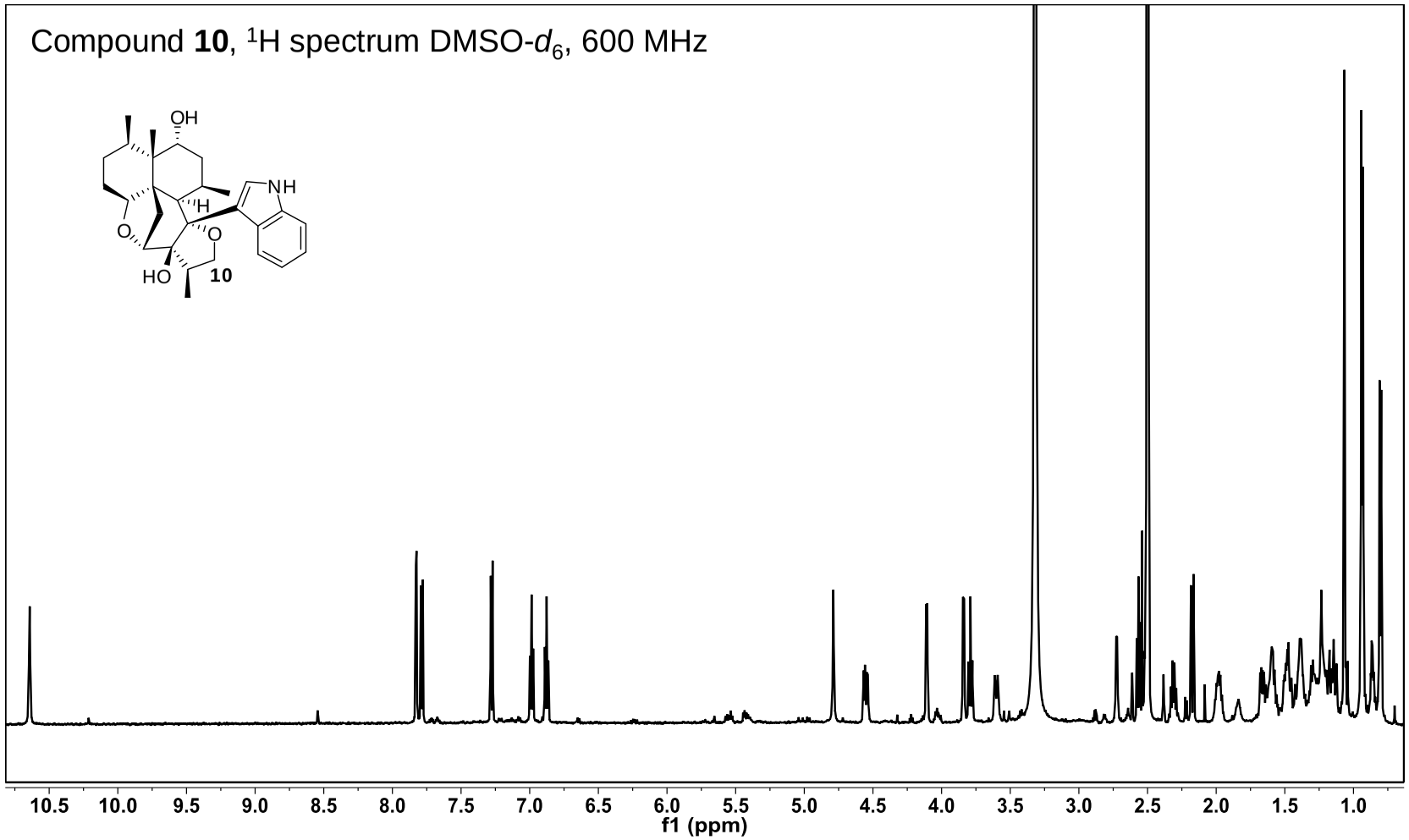
6

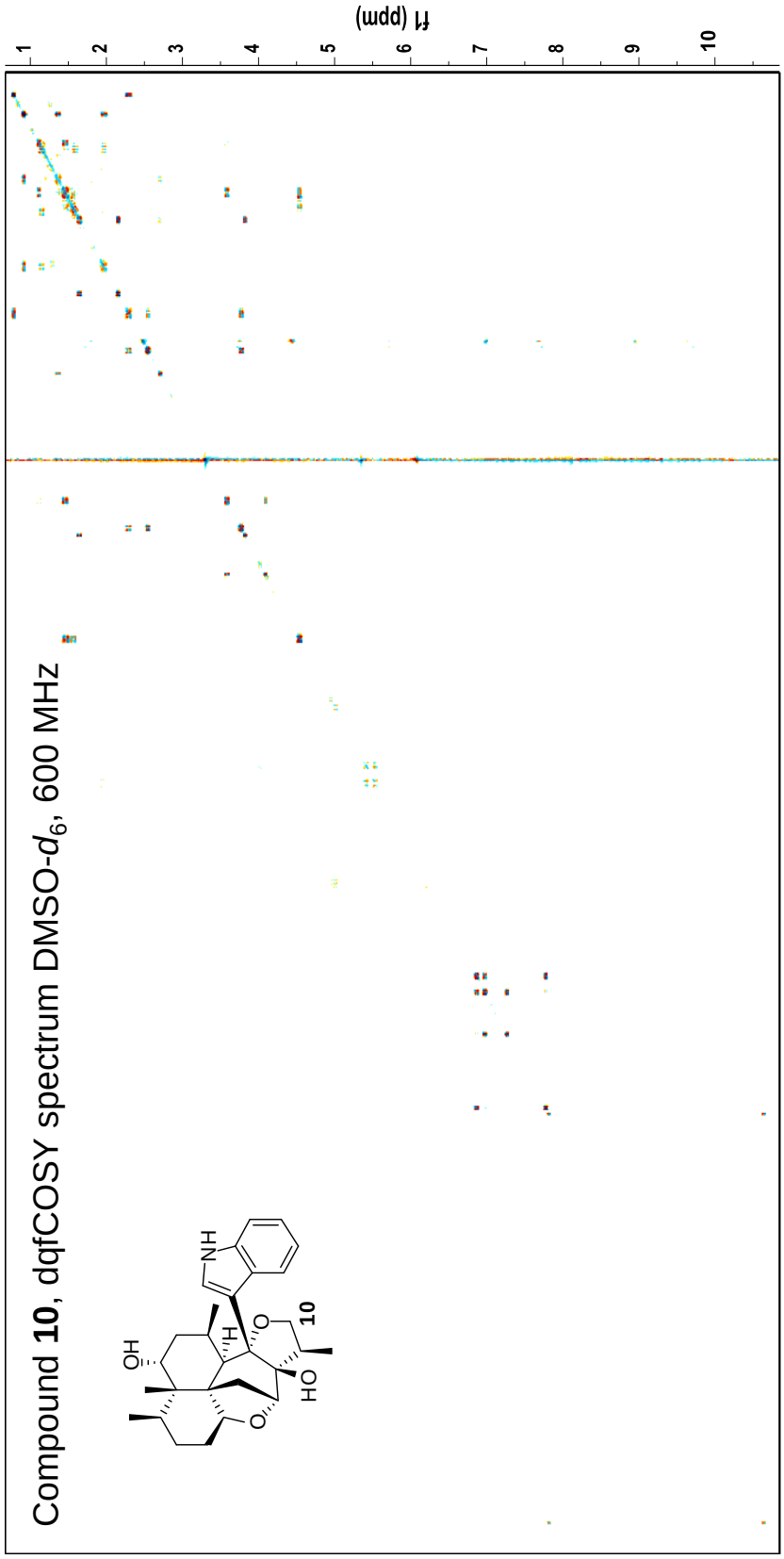

Compound **10**, ^{1}H spectrum methanol-d_4, 600 MHz
OH
NH
H
O
O
HO
10
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
f1 (ppm)

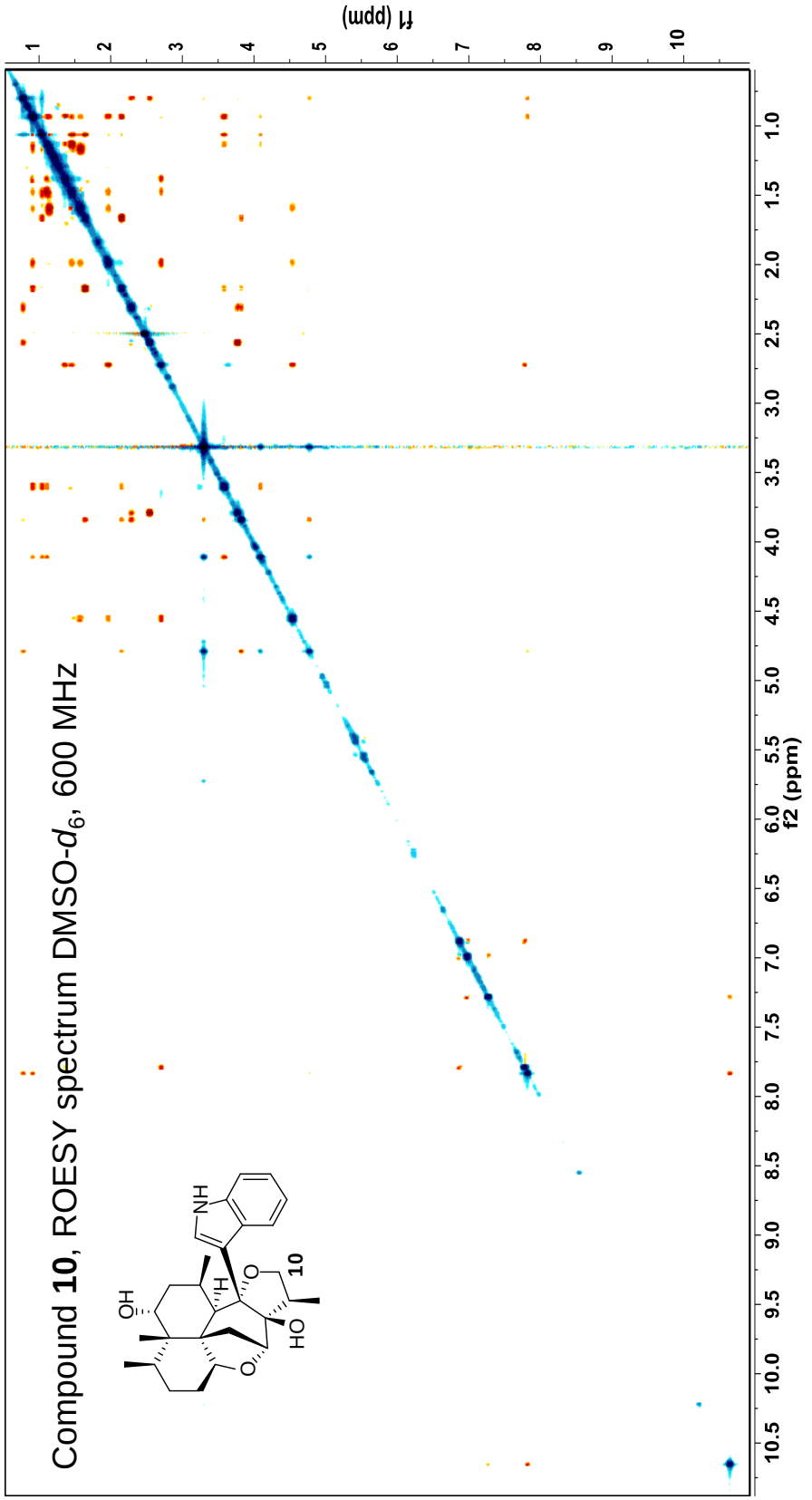

Compound **10**, dqfCOSY spectrum methanol-d_4, 600 MHz
f1 (ppm)
f2 (ppm)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
NH
OH
HO
O
O
H
10


Compound **10**, gHMBCAD spectrum methanol-d_4, 600 MHz
f1 (ppm)
f2 (ppm)
10
20
30
40
50
60
70
80
90
100
110
120
130
140
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
NH
OH
HO
O
O
H
10


Compound **10**, gHSQCAD spectrum methanol-d_4, 600 MHz
OH
NH
H
O
O
HO
10
f2 (ppm)
f1 (ppm)
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
10
20
30
40
50
60
70
80
90
100
110
120
130


Compound **10**, ^{1}H spectrum DMSO-d_6, 600 MHz
OH
NH
H
O
O
HO
10
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
f1 (ppm)


Compound 10, dqfCOSY spectrum DMSO-d_6, 600 MHz
f1 (ppm)
f2 (ppm)
NH
OH
HO
10


Compound **10**, ROESY spectrum DMSO-d_6, 600 MHz
f1 (ppm)
f2 (ppm)
1
2
3
4
5
6
7
8
9
10
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
NH
OH
HO
O
O
H
10

REFERENCES

(1) Szewczyk, E.; Nayak, T.; Oakley, C. E.; Edgerton, H.; Xiong, Y.; Taheri-Talesh, N.; Osmani, S. A.; Oakley, B. R. *Nat Protoc* **2006**, *1*, 3111.

(2) Amaike, S.; Keller, N. P. *Eukaryot Cell* **2009**, *8*, 1051.

(3) Miller, B. L.; Miller, K. Y.; Timberlake, W. E. *Mol Cell Biol* **1985**, *5*, 1714.

(4) McDonald, T.; Brown, D.; Keller, N. P.; Hammond, T. M. *Mol Plant Microbe Interact* **2005**, *18*, 539.

(5) Bok, J. W.; Chiang, Y. M.; Szewczyk, E.; Reyes-Dominguez, Y.; Davidson, A. D.; Sanchez, J. F.; Lo, H. C.; Watanabe, K.; Strauss, J.; Oakley, B. R.; Wang, C. C.; Keller, N. P. *Nat Chem Biol* **2009**, *5*, 462.

(6) Shimizu, K.; Keller, N. P. *Genetics* **2001**, *157*, 591.

(7) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *PNAS USA* **2009**, *106*, 7708.

(8) Campo, V. L.; Martins, M. B.; da Silva, C. H. T. P.; Carvalho, I. *Tetrahedron* **2009**, *65*, 5343.

(9) Van Lanen, S. G.; Dorrestein, P. C.; Christenson, S. D.; Liu, W.; Ju, J.; Kelleher, N. L.; Shen, B. *J Am Chem Soc* **2005**, *127*, 11594.

(10) Schneider, P.; Bouhired, S.; Hoffmeister, D. *Fungal Genet Biol* **2008**, *45*, 1487.

(11) Zhang, J.; Peng, Q.; Zhang, S.; Li, Y.; Li, S.; Gao, H.; Zhou, Z. *J Mol Str* **2011**, *987*, 34.

(12) He, Z.-M.; Price, M.; OBrian, G.; Georgianna, D. R.; Payne, G. *BMC Microbiol* **2007**, *7*, 104.